ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 31 (2006) 1487 – 1505

Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data

Stephen M. Smith,[a,*] Mark Jenkinson,[a] Heidi Johansen-Berg,[a] Daniel Rueckert,[b] Thomas E. Nichols,[c] Clare E. Mackay,[a] Kate E. Watkins,[a] Olga Ciccarelli,[d] M. Zaheer Cader,[a] Paul M. Matthews,[a] and Timothy E.J. Behrens[a]

[a]*Oxford University Centre for Functional MRI of the Brain (FMRIB), Dept. Clinical Neurology, University of Oxford, UK*
[b]*Department of Computing, Imperial College London, UK*
[c]*Department of Biostatistics, University of Michigan, UK*
[d]*Institute of Neurology, University College London, UK*

Received 10 October 2005; revised 17 February 2006; accepted 22 February 2006
Available online 19 April 2006

There has been much recent interest in using magnetic resonance diffusion imaging to provide information about anatomical connectivity in the brain, by measuring the anisotropic diffusion of water in white matter tracts. One of the measures most commonly derived from diffusion data is fractional anisotropy (FA), which quantifies how strongly directional the local tract structure is. Many imaging studies are starting to use FA images in voxelwise statistical analyses, in order to localise brain changes related to development, degeneration and disease. However, optimal analysis is compromised by the use of standard registration algorithms; there has not to date been a satisfactory solution to the question of how to align FA images from multiple subjects in a way that allows for valid conclusions to be drawn from the subsequent voxelwise analysis. Furthermore, the arbitrariness of the choice of spatial smoothing extent has not yet been resolved. In this paper, we present a new method that aims to solve these issues via (a) carefully tuned non-linear registration, followed by (b) projection onto an alignment-invariant tract representation (the "mean FA skeleton"). We refer to this new approach as Tract-Based Spatial Statistics (TBSS). TBSS aims to improve the sensitivity, objectivity and interpretability of analysis of multi-subject diffusion imaging studies. We describe TBSS in detail and present example TBSS results from several diffusion imaging studies.

Keywords: Diffusion imaging; DTI; Fractional anisotropy; FA; Morphometry; VBM

Introduction

The diffusion of water in brain tissue is affected by the local tissue microstructure; for example, it diffuses more easily along the major axis of a white matter fibre bundle than perpendicular to it (Moseley et al., 1990). Magnetic resonance diffusion tensor imaging (DTI) is sensitive to water diffusion characteristics (such as the principal diffusion direction and the diffusion anisotropy) and has therefore been developed as a tool for investigating the local properties of brain tissues such as white matter tracts (Le Bihan, 2003). There has also been a great deal of interest in using diffusion anisotropy as a marker for white matter tract integrity, for example, for disease diagnosis, tracking disease progression, finding disease sub-categories, studying normal development/aging, and as complementary information to investigating normal brain function (Horsfield and Jones, 2002; Lim and Helpern, 2002; Moseley, 2002; Neil et al., 2002; Pagani et al., 2005).

Diffusion anisotropy describes how variable the diffusion is in different directions and is most commonly quantified via a measure known as fractional anisotropy (FA) (Pierpaoli and Basser, 1996). It is highest in major white matter tracts (maximum theoretical value 1) and lower in grey matter while approaching 0 in cerebrospinal fluid. As a marker for tract integrity, FA is a useful quantity to compare across subjects as it is computable voxelwise and is a scalar value that is independent of the local fibre orientation (and therefore a relatively objective and straightforward measure to compare across subjects). Some researchers have simply summarised diffusion characteristics globally (for example, histogram-based summary measures of fractional anisotropy (Cercignani et al., 2001, 2003)), in order to compare different subjects. However, most recent work has been interested in spatially localising interesting diffusion-related changes. Many studies have, to this end, followed similar approaches to voxel-based morphometry (VBM, originally developed for finding local changes in grey matter density in T1-weighted structural brain images (Ashburner and Friston, 2000; Good et al., 2001)). In VBM-style FA analysis, each subject's FA image is registered into a standard space, and then voxelwise statistics are carried out to find areas which correlate to the covariate of interest (e.g., patients vs. normals, disability score, age).

* Corresponding author.
E-mail address: steve@fmrib.ox.ac.uk (S.M. Smith).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.02.024

There has been much debate about the strengths and limitations of VBM (Bookstein, 2001; Ashburner and Friston, 2001; Davatzikos, 2004; Ashburner and Friston, 2004). Some researchers continue to doubt the general interpretability of the results from this approach, primarily because there can be ambiguity as to whether apparent changes are really due to change in grey matter density or simply due to local misalignment, though it does seem that through careful application and validation, structural imaging studies using VBM can draw valid conclusions (e.g., Watkins et al., 2002). However, the potential problems with VBM-style approaches for data such as multi-subject FA images have not yet been investigated fully. In particular, this use raises a serious question, which has not yet been satisfactorily answered: how can one guarantee that any given standard space voxel contains data from the same part of the same white matter (WM) tract from each and every subject? In other words, how can we guarantee that registration of every subject's data to a common space has been totally successful, both in terms of resolving topological variabilities and in terms of the exact alignment of the very fine structures present in such data? A second problem relates to the standard practice of spatially smoothing data before computing voxelwise statistics—the amount of smoothing can greatly affect the final results, but there is no principled way of deciding how much smoothing is the "correct" amount (Jones et al., 2005). (Smoothing also increases effective partial voluming, a problem with VBM-style approaches particularly when applied to data such as FA; see Discussion for more comment on this.)

In this paper, we propose an approach to carrying out localised statistical testing of FA (and other diffusion-related) data that should alleviate the alignment problems. We project individual subjects' FA data into a common space in a way that is not dependent on perfect nonlinear registration. This is achieved through the use of (a) an initial approximate nonlinear registration, followed by (b) projection onto an alignment-invariant tract representation (the "mean FA skeleton"). No spatial smoothing is necessary in the image processing. We refer to this new approach as Tract-Based Spatial Statistics (TBSS). In the next section, we discuss, in slightly more depth, VBM-style approaches, and review some alternative approaches published to date. In following sections, we describe the proposed approach in detail, giving various example images illustrating the different analysis stages involved. Finally, we present example TBSS results from several DTI-based imaging studies.

Background: analysis of multi-subject diffusion data

VBM—overview and application to diffusion data

Voxel-based morphometry (Ashburner and Friston, 2000; Good et al., 2001) has been used in many structural imaging studies, looking for localised differences in grey matter density, typically between two groups of subjects. The common approach can be simply summarised:

- (Optional) Create a study-specific registration template by aligning all subjects' structural images to an existing standard space template image (such as the MNI152). Average all aligned images to create the new template, and optionally smooth.
- Align all subjects' structural images to the chosen template, normally first using affine and then low degrees-of-freedom nonlinear registration.
- Segment each subject's structural image into different tissue types. Generally use only the grey matter (GM) segmentation output.
- Smooth the segmentation output data. This is done for several reasons. First, smoothing of grey matter segmentation output produces an image which is intended to represent local "grey matter density"—i.e., producing a measure of the local balance between the count of GM and non-GM voxels. Second, the smoothing helps ameliorate the effects of misalignment of structures when the registration is imperfect. Third, it can increase sensitivity if the extent of smoothing matches the size of an effect of interest. Fourth, smoothing renders the data more Gaussian distributed, improving the validity of the commonly used Gaussian random field (GRF) theory thresholding approach. Typically between 4 and 16-mm full-width half maximum (FWHM) smoothing (with a Gaussian linear filter) is applied.
- Carry out voxelwise statistics, using any relevant covariates for the design matrix. A simple example would model group membership (patient and control), with appropriate contrasts comparing the group means. The standard approach is to use simple univariate statistics, meaning that each voxel is processed separately—the data for each voxel constitute a 1D vector of values, where that one dimension is subject number, and the model is fit separately to each voxel's 1D data vector.
- Threshold the resulting statistical T, F or Z image, taking into account multiple comparison correction. This is typically done using GRF (Worsley et al., 1992), using either a voxel-based or cluster-based approach (though extent-based thresholding can lead to false positives in VBM due to smoothness non-stationarity (Ashburner and Friston, 2000)).

There are also various optimisations (Good et al., 2001) that have been suggested to the above analysis protocol, such as using the GM segmentation to drive the registration (instead of the raw structural images), to make the registration better conditioned, and modulating the segmentation output after nonlinear registration, in order to compensate for local changes in volume caused by the alignment process.

VBM is most commonly carried out using the SPM software package (as an indication of this, all the references in the following paragraph used SPM), though the approach is sufficiently straightforward that several other freely available packages have also been used for "VBM-style" analyses. One of the reasons VBM has become popular is that it allows one, subject to interpretation caveats, to find changes anywhere in the brain—it is not necessary to prespecify regions or features of interest.

Recently, researchers have applied VBM-style analysis to test for localised changes in diffusion-related images. Most commonly, this has involved testing FA images for voxelwise differences between two groups of subjects. The registration is performed either using structural images or by using the FA images directly. No segmentation step is necessary. Smoothing is usually carried out (with no general agreement on how much is appropriate) before running standard voxelwise statistics and thresholding. Typical examples of this kind of approach can be found in Simon et al. (2005), studying chromosome 22q11.2 deletion syndrome, using

12-mm FWHM smoothing (Rugg-Gunn et al., 2001; Eriksson et al., 2001), studying epilepsy, using 8-mm FWHM smoothing (Barnea-Goraly et al., 2003), studying fragile X syndrome, using 4-mm FWHM smoothing and (Büchel et al., 2004), testing for L-R asymmetry and handedness, using both 4- and 12-mm FWHM smoothing.

Alignment issues in VBM-style analyses

Various papers (Bookstein, 2001; Ashburner and Friston, 2001; Davatzikos, 2004; Ashburner and Friston, 2004) have discussed the limitations and strengths of VBM-style approaches. It has been observed in particular that one must be very careful not to misinterpret residual misalignments. How can one guarantee that any given voxel (in the final space in which voxelwise statistics will be carried out) contains data from anatomically corresponding regions—i.e., the same part of the same white matter tract from each and every subject? In the context of VBM-style analysis of FA data, consider the following scenario: a patient group includes individuals with greater ventricular sizes than a control group. The two groups, however, have the same basic white matter integrity. Because of the differences in ventricular configuration, conventional (low to medium degrees-of-freedom) registration approaches will shift the anterior section of the corpus callosum (CC) anteriorly in the patient group relative to the controls; registration of the data (and subsequent smoothing) is unlikely to fully remove this group difference in alignment. When voxelwise statistics are carried out, this residual misalignment shows up as a group difference in FA; at the front of the CC, it appears that FA(patients) > FA(controls), while at the back, the reverse is implied.

This problem is discussed further in Simon et al. (2005), where the authors are careful to interpret apparent FA changes as being in fact due to changes in ventricle size. A further example of this danger can be seen in Vangberg et al. (2005), where the results are strongly suggestive of a shift of the pyramidal tract, rather than a true change in WM integrity.

Some researchers, aware of this problem, use careful post-stats analyses to help disambiguate the interpretation of apparent differences. For example, Sommer et al. (2002) used a standard VBM-style approach (using 6-mm FWHM smoothing) and then checked afterwards that the alignment was reasonable, looking at the WM-masked region-of-interest (ROI) in the unsmoothed FA images, near the reported difference. However, the reported FA difference is very close to cortical grey matter, and it is difficult to be sure that differences in GM/WM partial volume effects have not contributed to the result.

There have been various papers presenting investigations of alignment issues specific to diffusion tensor data. (Jones et al., 2002) use FA to drive affine alignment across subjects. Park et al. (2003) investigates alignment when driven by a variety of diffusion-derived measures; high degrees-of-freedom nonlinear registration is used, and alignment success is quantified via similarity of final tractography maps. It is shown that using all 6 tensor components to drive the registration similarity cost function gives better overall alignment than other combinations of DTI-derived information, including FA (although the differences were not large). In Park et al. (2004), this approach was then used to compare white matter structure in schizophrenics relative to controls. Their registration does appear to help with the alignment issues discussed above, but, even with this relatively sophisticated registration approach, the authors state that there were still "some registration errors in the boundary of narrow fibre bundles" and, for this reason, did not directly compare their VBM-based asymmetry tests between schizophrenics and controls.

It would appear that in general, it is not safe to assume that (even high degrees-of-freedom nonlinear) registration can align FA data sufficiently well across subjects to allow simple unambiguous interpretation of voxelwise statistics. Also, if one cannot guarantee that alignment is "correct", then it must be assumed that sensitivity to true differences is suboptimal.

The registration problem is not resolved even if one takes the degrees-of-freedom to the extreme, forcing all images to look extremely similar (this is an option with some nonlinear registration approaches); although it may be possible to distort one image to look very much like another, one does not necessarily have confidence that any given structure has in fact been aligned to that same structure in the other subject. Some nonlinear registration methods are able to go so far in making one image look like another that they can even break the "topology" of the image being distorted—for example, a single fibre bundle may be split into two disconnected bundles, or two distinct tracts could be merged into one.

Smoothing issues in VBM-style analyses

A second problem with VBM-style analyses lies in the arbitrary choice of smoothing extent. Smoothing can help ameliorate residual misalignments, though not in a well-controlled way. It can also help improve sensitivity in the detection of changes, if the extent of smoothing is matched to the spatial extent of the structure of interest. However, it is not generally known in advance what this will be, so there is no principled way to choose the smoothing extent. If one were to try a range of smoothing extents, the final interpretation can become more confused, and multiple-comparison corrections need to be made more aggressive.

This issue is investigated in detail in Jones et al. (2005), where it is shown that the final results (of VBM-style FA analysis of schizophrenia data) depend very strongly on the amount of smoothing. Different smoothing extents (from 0- to 16-mm FWHM) are applied, and apparent group differences appear and disappear across the different tests. Likewise, Park et al. (2004) also investigated asymmetry in schizophrenia, using 3-, 6-, and 9-mm FWHM smoothing; several of the apparent asymmetries were quite different in the different cases.

As well as the problem of the arbitrariness of choice of smoothing extent, smoothing increases the partial voluming problem; one would like to know whether any estimated change in FA is due to a change in FA in white matter rather than a change in the relative amounts of different tissue types, but smoothing exacerbates this ambiguity. If possible, it would be good to obviate the need for spatial smoothing of diffusion data in such applications.

Alternative strategies for localising change

A simple alternative to VBM-style FA analysis is to specify an ROI, usually carried out in practice by hand, separately for each subject (e.g., Ellis et al., 1999; Kubicki et al., 2003). FA values are taken from the ROI(s) and then compared across subjects. In the centres of the largest tracts this can be a reliable approach; however, it can be hard to objectively place ROIs for smaller/

thinner tracts, particularly given partial volume issues. Furthermore, this kind of approach limits a study to only being sensitive to change in those few parts of the brain where ROIs are placed. See references in Park et al. (2003) for more examples of this kind of approach.

More sophisticated approaches use tractography (fibre bundle tracking, e.g., Conturo et al., 1999; Behrens et al., 2003a) to identify voxels from which to take FA values for cross-subject comparison. In such approaches, the relevant tracts are usually identified by initialising/constraining tractography using hand-drawn ROIs. For example, in Pagani et al. (2005), DTI-related changes in the pyramidal tracts were observed in patients with early MS-like symptoms. ROIs were hand-drawn in a standard space to identify the pyramidal tracts. These were then used to seed streamlining-based tractography in each subject's original DTI data, to define in each the pyramidal tract. The results were then averaged to provide a mean pyramidal tract mask. Tests were then carried out on various DTI-related metrics by affine-aligning patient data into MNI152 space and taking summary statistics using all voxels within this mask.

In the above approach, tractography is used to determine a standard space ROI, but the final analysis still depends critically on the accuracy of alignment of each subject to the standard space. In Jones et al. (2006), this problem is avoided by using each subject's tractography results to estimate mean FA in several major tracts, summarising each tract with a single mean FA value before comparing normals and schizophrenics.

A still more sophisticated approach is to compare the variation of FA values along the tractography-derived fibre bundles directly across subjects, by first parameterising the space of the fibre bundle, e.g., according to distance along the bundle. This does not then rely strongly on perfect cross-subject alignment. (In our method described below, we attempt to combine the strength of this kind of approach with the investigative power and ease-of-use of voxelwise analyses.) An example can be found in Gong et al. (2005). Tractography is used to find cingulum bundles, and FA is parameterised according to the position within a tract. This allows cross-subject comparison of FA values along the given tract without requiring accurate final registration. In a similar approach, Gerig et al. (2005) finds tract bundles based on an initial hand-drawn ROI, and then parameterises FA (and other DTI-derived measures such as ADC and tensor eigenvalues) along the resulting bundles. For a given subject scanned on 6 occasions, all measures are shown to be reproducible (at one point on the bundle) to between 5 and 10%.

A limitation of such approaches is that only those tracts that can be reliably traced (and separated from other tracts) can be used to create relevant FA parameterisation. As there is not at present a robust, fully automated, way of finding and classifying all brain tracts, only those tracts that have been specifically analysed (usually using hand-drawn ROIs and various termination heuristics) can be investigated. A second problem is that it may not be straightforward to objectively and accurately identify the effective ends of tracts of interest, creating possible problems for parameterisation that is objectively consistent across subjects. A third limitation relates to partial volume effects at the edges of the tracts. By definition, the fibre bundle "edges" (as found by tractography) contain some non-bundle partial volume fraction; in general, the amount of non-tract partial volume included in the FA parameterisation is not well controlled, causing some arbitrariness in the final sampled FA values when using certain ways of measuring FA, such as mean value across the tract cross-section.

Method: tract-based spatial statistics

Overview of TBSS

As discussed above, strengths of VBM-style analyses are that they are fully automated, simple to apply, investigate the whole brain, and do not require prespecifying and prelocalising regions or features of interest. Limitations include problems caused by alignment inaccuracies, and the lack of a principled way for choosing smoothing extent. Tractography-based approaches have fairly complementary advantages and disadvantages. They can overcome alignment problems by working in the space of individual subjects' tractography results and for similar reasons do not necessarily require presmoothing. However, such approaches do not allow the whole brain to be investigated and generally require user intervention in order to define the tracts to be used.

In TBSS, we attempt to bring together the strengths of each approach. We aim to solve the alignment and smoothing issues, while being fully automated, investigating the "whole" brain—not requiring prespecification of tracts of interest. This is achieved by estimating a "group mean FA skeleton", which represents the centres of all fibre bundles[1] that are generally common to the subjects involved in a study. Each subject's FA data is then projected onto the mean FA skeleton in such a way that each skeleton voxel takes the FA value from the *local centre* of the nearest relevant tract, thus hopefully resolving issues of alignment and correspondence. To briefly summarise the TBSS approach:

- Identify a common registration target and align all subjects' FA images to this target using nonlinear registration. At this stage, perfect alignment is not expected or required.
- Create the mean of all aligned FA images and apply "thinning" (non-maximum-suppression perpendicular to the local tract structure), to create a skeletonised mean FA image. Threshold this to suppress areas of low mean FA and/or high inter-subject variability.
- Project each subject's (aligned) FA image onto the skeleton, by filling the skeleton with FA values from the nearest relevant tract centre. This is achieved, for each skeleton voxel, by searching perpendicular to the local skeleton structure for the maximum value in the subject's FA image.
- Carry out voxelwise statistics across subjects on the skeleton-space FA data.

We now describe each step in more detail.

Preprocessing

A single diffusion dataset typically comprises between 7 and 200 separate 3D images; these encode diffusion strength in various different directions, as well as including one or more images with

[1] "Fibre bundle" is usually taken to mean a collection of white matter neurons all following a similar anatomical path (at least locally), while "tract" is sometimes used to mean individual axons, but more commonly to fibre bundles. In this paper, we generally intend the latter and therefore use the terms "tract" and "fibre bundle" interchangeably.

no diffusion weighting. A common preprocessing step is to align all the images with each other before estimating diffusion-related measures such as the diffusion tensor, principal diffusion direction, and fractional anisotropy. This prealignment (similar to motion correction in fMRI data) is both to correct for head motion during the session and to reduce the effects of gradient coil eddy currents (Horsfield, 1999). While head motion mostly causes rigid-body image motion, eddy currents appear as a (slightly more general) linear image transformation, to first order. We therefore use FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002) to apply full affine (linear) alignment of each image to the no-diffusion-weighting image, using the mutual information cost function.

After data prealignment, the diffusion tensor can be calculated, normally using a simple least squares fit of the tensor model to the diffusion data. From this, the tensor eigenvalues can be calculated, describing the diffusion strength in the primary, secondary and tertiary diffusion directions. From these, it is straightforward to calculate FA (Basser et al., 1994; Pierpaoli and Basser, 1996).

Finally, we apply BET (Smith, 2002) brain extraction to the non-diffusion-weighted image, to exclude non-brain voxels from further consideration.

Nonlinear alignment

The first step in aligning multiple FA images to each other is a voxelwise nonlinear registration, driven by the FA images themselves. We do not want to change the fundamental nature of the images during this alignment – we want to keep the general tract structure intact – but we need to align the images sufficiently well that the second stage (projection of data onto a tract skeleton) functions correctly. We therefore need nonlinear alignment having *intermediate* degrees of freedom (DoF).

At the low-DoF extreme (for example, affine-only registration with no nonlinear component), there is sufficiently little guarantee of alignment of even the most major tracts, that voxelwise statistics across subjects is unwise.

At the high-DoF extreme (high-dimensional warping), it is possible to align two images almost perfectly, so that they look almost exactly like each other; the problem here is that in order to achieve this, the original images have been warped so much that one may not have preserved the overall structure, i.e., how the different features (in this case, different white matter tracts) relate to each other. A given tract (e.g., cingulum bundle) may be warped so far that it becomes aligned to a totally different tract in the target image (e.g., corpus callosum). Furthermore, the warp may be "non-homologous" – image topology may be changed – for example, two tracts may be merged into one or one tract may be split into two. In summary, current high-DoF methods cannot be considered to produce reliable homologies.

We want to avoid either extreme—it is important to align subjects' data together to make local comparison possible, but with some restriction applied to the warp, so that the overall structure topology is preserved. To this end, we use a generic nonlinear registration method which is capable of high-dimensional warping but which can also be robustly controlled to limit the effective dimensionality, to give us the desired restriction on warp complexity. We have chosen to use a nonlinear registration approach based on free-form deformations and B-Splines (Rueckert et al., 1999) which is available from www.doc.ic.ac.uk/~dr as a package called the "Image Registration Toolkit" (IRTK). The aim of free-form deformations is to deform an image by moving the control points of an underlying mesh. The warp field applied is found for image positions between the mesh control points using B-spline interpolation. The optimal warp is found by moving the control point locations until the registration cost function is minimised. This cost function attempts to both optimise a voxel-based similarity measure at the same time as imposing regularisation (smoothness) on the warp field.

For this application, we have used cross-correlation for the similarity measure, as an inter-modal cost function is not needed when aligning different FA images together. We set a control point spacing of 20 mm and set additional regularisation to zero[2]. Thus, the smoothness of the warp field is determined purely by the control point spacing, which here is chosen to be large enough to achieve what is considered to be an appropriate degree of warp complexity, as discussed above. The nonlinear registration is preceded by affine-only registration, to achieve initial alignment. Running IRTK with these options takes approximately 20 min on a modern desktop computer, to align a single FA image to a different FA target.

Identifying the target for alignment

Upon investigation of the quality of registrations obtained by applying IRTK to typical FA images (typical resolutions being between $2 \times 2 \times 2$ mm^3 and $4 \times 4 \times 4$ mm^3), it was found that registration is more successful if the target is a *real* FA image rather than a (blurred) average FA image. This is perhaps unsurprising, as a single subject will be sharper than an averaged image, giving "better" information to drive the alignment, as long as topology is sufficiently similar to the input image.

We therefore identify a single subject's FA image to act as the target for all nonlinear registrations. We want this subject to be the "most typical" subject of the entire group, i.e., to be the target image which minimises the amount of warping required for all other subjects to align to it. To find this most typical subject, we register every subject to every other subject, summarise each warp field by its mean displacement, and choose the target subject as being the one with the minimum mean distance to all other subjects. Because the affine part of these registrations is robust and does not contain any interesting information about a subject's tract topography/topology (in this context), the effect of the initial affine transformation is subtracted from the estimation of a warp field's mean displacement distance.

An alternative, faster, approach, would be to choose an initial target at random, register every subject to this, and use warp field concatenations to estimate the above. However, given the complex, multidimensional search strategy involved in finding an optimal warp between two images, and given possible topology changes between subjects, it is safer to take the full search strategy described above. We tested whether it was more robust to summarise a warp field with the median displacement instead of the mean, but this made no difference to the choice of optimal

[2] This spacing was primarily optimised empirically but also relates to the scale of the typical spacing of different parts of the final skeleton, and hence the amount of movement needed to prealign FA images, as well as the maximum search distance needed to project each subject's FA onto the skeleton—see later for further description of these aspects of the approach.

target in all 10 studies tested. See also (Kochunov et al., 2005; Guimond et al., 2000) for further discussion of mean atlas spaces.

Creating the mean FA image and its skeleton

After identifying the most typical subject as the target, all subjects' FA images are aligned to this, and then the entire aligned dataset is affine-transformed into $1 \times 1 \times 1$ mm^3 MNI152 space; all subsequent processing is carried out using this space and resolution. The choice of MNI152 space is made for convenience of interpretation and display. The choice of a higher resolution here than typical diffusion datasets means that there is no significant interpolation blurring (i.e., increase in partial voluming) when the nonlinear warp plus standard-space affine transformation is applied to each individual subject's data. Using an even higher resolution than this would bring increasingly little benefit, but simply result in slower computation and unnecessarily large data files. Note, however, that if and when higher resolution diffusion data is acquired, it will be straightforward to increase the working resolution for the post-alignment steps.

The transformed FA images are now averaged to create a mean FA image. This image is locally relatively smooth, both because of the effect of averaging FA images across subjects, and because of the resolution upsampling. Fig. 1 (top-left) shows an example axial slice through a mean FA image.

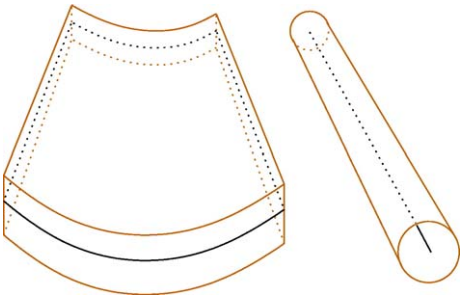

Fig. 2. Examples of fibre bundles; a "thick sheet" with a thin surface as its skeleton, and a "tube", with a line as its skeleton.

The mean FA is now fed into the tract skeleton generation, which aims to represent all tracts which are "common" to all subjects. The skeleton will represent each such tract as a single line (or surface) running down the centre of the tract. Most contiguous sets of tracts appear topologically to be curved sheets of a certain thickness (e.g., corpus callosum), or, less frequently, curved "tubes" (e.g., the cingulum bundle); see Fig. 2. In the former case, we want the skeleton to be a thin curved surface running down the centre of the sheet, and in the latter, we want the skeleton to be a curved line running down the centre of the tube. Away from the centre surface or line, the FA values fall off gradually, becoming very low as one

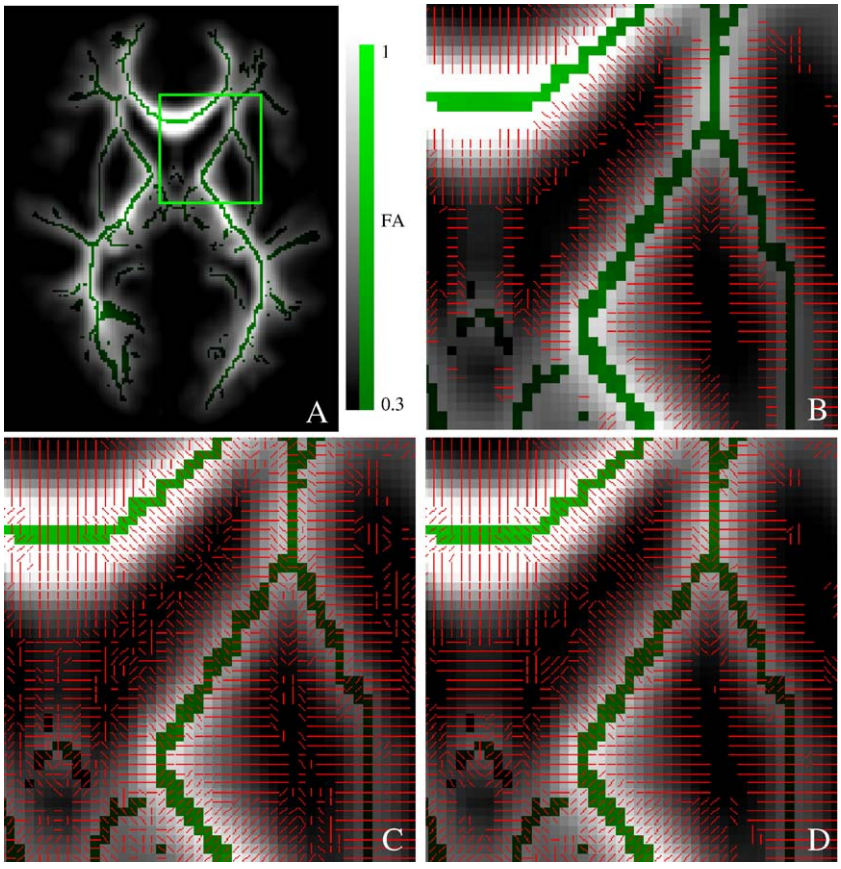

Fig. 1. Different skeletonisation stages. (A) Original mean FA image with final skeleton and the ROI used for the remaining sub-images. (B) Skeletonisation stage 1, using local FA centre-of-gravity to find tract perpendiculars. (C) Skeletonisation after stage 2, using FA image second-derivative to find remaining perpendiculars. (D) Result of smoothing the perpendicular direction vector image. Note that the tract appears more than a single voxel thick in some places, because of its 3D nature; where the fibre bundle surface lies partially parallel to the plane being viewed, it will not appear thin, though would do if viewed with a different 3D slicing.

moves out of white matter. To achieve skeletonisation we first estimate the local surface perpendicular direction (at all voxels in the image) and then perform non-maximum-suppression in this direction. In other words, a search is made along all voxels in the local "tract perpendicular direction", and the voxel with the highest FA is identified as the centre of the tract.[3]

The local tract surface orientation is found as follows. If the voxel of interest lies away from a tract centre, FA will be higher in the neighbouring voxels on one side of the voxel than on the other—the direction in which it is highest points towards the nearest tract centre. We quantitate this by finding the centre-of-gravity of the local $3 \times 3 \times 3$ voxel neighbourhood (effectively we are taking the first derivative of the FA image). The vector from the current voxel centre to the local centre-of-gravity (CofG, of FA values) should point towards the tract centre, in a direction perpendicular to the local tract structure. Therefore, as long as the local CofG does not lie close (within 0.1 mm) to the centre of the current voxel, the perpendicular direction is assumed to be given by this vector. See Fig. 3 for an example.

Alternatively, if the local CofG does lie close to the centre of the current voxel, it is assumed that one is very near to the tract centre, and an alternative method of estimating the perpendicular is used. In this case, the direction of maximum change is found; from the local $3 \times 3 \times 3$ voxel neighbourhood, the mean of each opposing pair of voxels is subtracted from the centre value, and the direction which causes this difference to be maximised is assumed to be perpendicular to the local tract (effectively we are taking the second derivative of the FA image).

Finally, we regularise the estimated tract perpendicular direction in order to improve estimation robustness; we replace each direction estimate with the mode of the quantised local $3 \times 3 \times 3$ set of estimated directions.

We are now in a position to search for the centre of each tract, i.e., form the tract skeleton. At each voxel we compare the FA value with the two closest neighbours on each side, in the direction of the tract perpendicular. If the FA value is greater than the neighbouring values, then the voxel is marked as lying on the skeleton.

Fig. 1 illustrates the various steps involved in turning a mean FA image into an FA tract skeleton. The top-left image shows an example axial slice through a mean FA image; overlaid is the final skeleton, and the ROI used for the remaining sub-images is shown. In top-right are the results of the first stage of estimation of the perpendicular direction to the local tract structure; the lines show the directions estimated on the basis of the local FA centre-of-gravity. Note that these are only estimable away from the tract centres. In bottom-left are the results after the second stage; where centre-of-gravity has not estimated the tract perpendicular, the FA image second derivative is used. Thus the local perpendicular direction is now estimated at all voxels where FA is not very close to zero. In bottom-right the direction estimates have been smoothed by taking the mode of the directions in the $3 \times 3 \times 3$ neighbourhood.

We now have an FA skeleton which should represent the different tract structures in the mean FA image. This is thresholded in order to restrict further analysis to points which are within white matter which has been successfully aligned across subjects. We have found that thresholding the mean FA value between 0.2 and 0.3 successfully excludes voxels which are primarily grey matter or CSF in the majority of subjects and also means that the skeleton does not run right up to the outermost edges of the cortex, where the constraints on the nonlinear alignment mean that the most variable (across subjects) tracts are not well aligned. In other words, we are excluding from further analysis those parts of the brain where we do not believe that we can assume good tract correspondence across subjects.

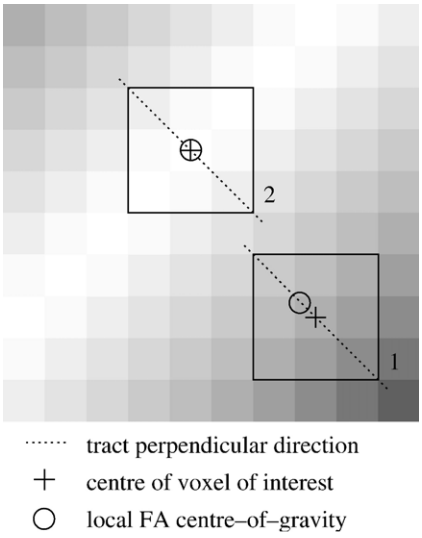

Fig. 3. Example of (1) a voxel where the local centre-of-gravity (CofG) points in the local tract perpendicular direction, and (2) a voxel lying on the tract centre.

Note that the skeleton tends to be disconnected at many junctions; this is primarily due to the fact that the tract perpendicular direction is not well-defined at junctions, and hence the non-maximum suppression "perpendicular" to the tract cannot function well. One could attempt to force connectivity at junctions, for example, through standard morphological processing, but this would probably be dangerous; the next stage, where FA data get projected onto the skeleton, would also not be well conditioned at junctions (for the same reason—i.e., lack of a well-defined projection direction), unless a much more sophisticated projection method was developed specifically for junctions.

Projecting individual subjects' FA onto the skeleton

We now "project" each subject's aligned FA image onto the mean FA skeleton. The aim here is to account for residual misalignments between subjects after the initial nonlinear registrations. At each point in the skeleton, we search a given subject's FA image in the (already-computed) perpendicular tract direction to find the maximum FA value and assign this value to the skeleton voxel. This effectively achieves alignment between the skeleton and this subject's FA image without needing perfect nonlinear preregistration. Any systematic difference in exact tract location between groups of subjects will thus not bias the comparison of FA values between the groups.

Note that this approach is effectively achieving fine alignment across subjects in the tract perpendicular, not in the direction parallel to the tract. This is what we require; FA changes very quickly as one moves perpendicular to the local fibre bundle, so even the smallest misalignments in this direction have great effect on the final FA statistics. Parallel to the tract, FA changes relatively

[3] The skeleton generation is probably not strongly dependent on the exact image processing method used here—for example, other thinning methods such as finding crest lines or medial axes would probably give similar results.

slowly, such that the alignment provided by the initial nonlinear registration is sufficient to align "like with like" across subjects.

There are two limits placed on this perpendicular search within a given subject's FA image. The first is that we constrain the search to remain closer to the starting section of skeleton than to any other section of skeleton; where two separate sections of the skeleton lie close to each other, the space in between is divided into two, and each skeleton section can only search voxels within its part of that space. This is achieved by creating a skeleton distance map—all voxels in the image are filled with a value encoding the distance to the nearest skeleton point. The above rule is then enforced by only searching outwards from a given skeleton point while this distance measure is increasing. Thus any given voxel can only be mapped into a single section of skeleton. Fig. 4 shows an example "distance map". The red–yellow overlay encodes, for each brain voxel, how far the nearest skeleton voxel is.

Secondly, there is a further constraint placed on the maximum search distance via a soft distance limit. A wide Gaussian function (FWHM 20 mm) is applied as a multiplicative weighting to FA values when carrying out the search for maximum FA (note—this is a weighting function in the search, not a smoothing). This deweights the most distant voxels in a smooth, controlled manner. Once the optimal voxel has been found, its FA value (not weighted by the distance function) is placed into the current skeleton voxel.

There is one major tract in the brain where the local skeleton topology is tubular rather than sheet-like—in the inferior part of the cingulum. The superior part of the cingulum (i.e., above the corpus callosum) is slightly extended across its cross-section in the inferior-superior direction, and well-localised across subjects by virtue of the strong, nearby corpus callosum, and hence the normal projections described above work well (similar issues relate to the fornix). In contradistinction, the inferior (retro-/infrasplenial) part of the cingulum is more tubular than sheet-like, and its position in any given axial slice varies across subject in both the anterior-posterior and left-right directions. Because of this, there is no well-defined search direction for the FA projection onto the skeleton, so we use a different approach here. The inferior cingulum is automatically defined via a liberal standard-space mask, and for skeleton points within this mask, the local search for maximum FA is within a circular space in the appropriate axial slice, rather than along a single "perpendicular" direction.

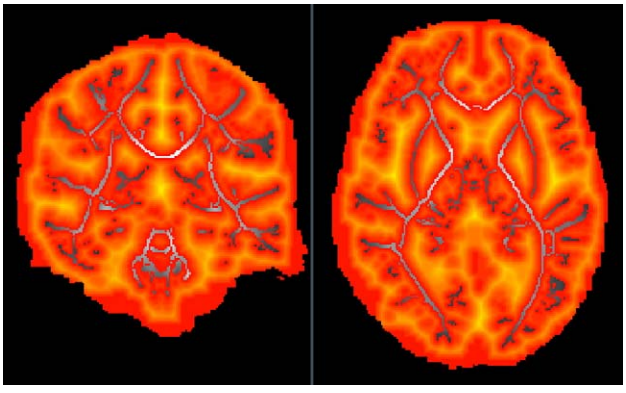

Fig. 4. Example "distance map": the red–yellow overlay encodes, for each brain voxel, how far the nearest skeleton voxel is. This is used during the projection of individual FA maps onto the skeleton in order to ensure that values are only taken onto the nearest part of the skeleton. The underlying mean FA skeleton can be seen where distance is zero.

We have therefore, for each subject, filled the skeleton with FA values from the centres of the nearest relevant tracts. Note that the idea of taking a "pure" FA value from the centre of a tract in a way that claims to be unaffected by partial volume effects is only strictly true for tracts wider than the relevant voxel dimension. When this is not the case, i.e., for the thinnest tracts, the "centre" peak FA value will reflect both the tract width and the true peak FA value, due to partial voluming.

Statistics and thresholding

At this point, we now have the data ready to feed into voxelwise cross-subject statistical analysis. Each subject's FA image has been prealigned to a common space using constrained nonlinear registration, a common tract skeleton has been formed, and each subject's FA image has then been fully aligned (via perpendicular search for local tract centre) with the common skeleton. Thus, the data are now in the form of a sparse (skeletonised) 4D image, with the fourth dimension being subject ID. We can now carry out voxelwise statistics across subjects, for each point on the common skeleton.

The simplest approach is to use univariate linear modelling, i.e., process each skeleton voxel independently, applying the general linear model (GLM, i.e., multiple regression) across subjects. For example, one can easily use a two-regressor analysis (equivalent to an unpaired t test) to test for significant local FA differences between a group of patients and a group of controls.

For simple parametric regression and inference to be valid, the cross-subject null distribution of FA values (for any given voxel) needs to be Gaussian. If we have succeeded, for any given skeleton voxel, in taking FA values from the centre of the same point of the same tract in all subjects, one would indeed expect Gaussian variability, except possibly for very high or very low mean FA values. In the Results section, we show some results of testing data Gaussianity; it is found that the TBSS-produced data are indeed Gaussian.

The remaining complication in carrying out inference is the issue of multiple-comparison correction. One would not want to apply Bonferroni correction, as the data will contain some intrinsic spatial smoothness (typically the final skeletonised FA data have intrinsic smoothness of order 4 mm FWHM), and this would therefore be an over-conservative correction. Because of the highly nonlinear steps leading to the formation of the skeletonised data, the lack of connectivity at many junctions, and the topological skeleton complexity, one also cannot assume the validity here of standard Gaussian random field theory (GRF, Worsley et al., 1992), unlike with standard VBM-style approaches; however, it may well be that approaches such as the application of GRF to 2D meshes containing MEG-derived data (Pantazis et al., 2005), or other probability validation work (PVW), could help here.

Alternatively, one could use a permutation-based approach (Nichols and Holmes, 2001), testing an appropriate test statistic (e.g., voxel t value, cluster size[4] or cluster mass) against the null distribution (generated via multiple random permutations of

[4] Note that for cluster-based inference, one needs to choose an initial cluster-forming threshold; the choice of this initial threshold is totally arbitrary, which is a limitation of current cluster-based approaches in general. However, note that the final ("corrected") P value associated with a cluster through permutation testing is totally valid, regardless of what cluster-forming threshold is used.

study	#	# controls	field (T)	resolution (mm)	directions x repeats	# b0	gradient direction scheme (b)	TR / TE (s / ms)
ALS	13	20	1.5	2.3x2.3x2.3	54 x 1	6	J-54 (1150) + J-6 (300)	4 / 106
schizophrenia	33	36	1.5	2.5x2.5x2.5	12 x 13	13	dodecahedral (1000)	5.4 / 76
MS	15	-	1.5	2.5x2.5x2.5	60 x 2	10	J-60 (1000)	8.5 / 89
stuttering	15	11	1.5	2.5x2.5x2.5	60 x 2	10	J-60 (1000)	8.5 / 89
normals (Fig.12)	-	18	1.5	2x2x2	60 x 3	15	J-60 (1000)	9.0 / 89
normals (repeatability)	-	8	1.5	2x2x2	60 x 3	15	J-60 (1000)	9.0 / 89

Fig. 5. Diffusion data acquisition protocols. The J-X gradient direction schemes create multiple directions equally spaced over a sphere, according to Jones et al. (1999).

subject ID ordering with respect to the model) of maximum (across space) values of the test statistic. This gives strong control of "familywise errors" while searching over the entire skeleton for regions of significant effect. This approach does not require the cross-subject distribution of FA values to be Gaussian.

Note that a general advantage of the skeleton-based approach is the reduction in the number of tests; fewer tests mean a less severe multiple comparisons problem.

Results

In the following sections, we present example results and quantitations from different stages of the TBSS analysis, followed by example results from several diffusion imaging studies. The data generally used to illustrate TBSS are taken from a study of amyotrophic lateral sclerosis (ALS, a progressive neurodegenerative disease most prominently affecting the motor system). The diffusion acquisition parameters for this and all other data used in this paper are given in Fig. 5.

Nonlinear alignment

In Fig. 6, we show example registrations of 3 controls and 3 ALS patients, with ROIs showing the corpus callosum. In each, the images on the left show affine-only registration, and on the right the full nonlinear registration results. In these examples, it is clear that affine-only registration is insufficient to give good alignment. The overlaid red edges are intensity edges from the target image. Further examples can be seen later in Fig. 11; the nonlinear registration is generally working well, but on close inspection, it is clear that "perfect" alignment has not been achieved, showing the insufficiency of pure nonlinear registration before applying voxelwise statistics.

Identifying the target for alignment

Fig. 7 shows example results of summary nonlinear displacement scores. The subjects are 20 controls and 13 ALS patients, respectively. For each target subject, a column of scores is shown; each score represents the root mean square displacement (across all brain voxels) for the nonlinear component of the alignment of any given subject to the target subject.

The diagonal is full of zeros as each subject does not need to deform to match itself. More interestingly, the matrix is fairly close to being symmetric (about the diagonal). This reflects the fact that in general, registering subject A to subject B involves a similar amount of deformation than registering B to A, as one would hope.

In the bottom row, each target subject's overall score is found by taking the mean of the scores from registering each other subject to the one in question. Note the relatively high variation in these mean scores, reflecting the fact that some subjects are significantly more "typical" to the group of subjects in question than others. Note also the greater variability within the patient group than within the control group. The means of the two groups, however, are not significantly different.

Creating the FA skeleton

Fig. 8 shows several orthogonal slices illustrating the mean FA image (red–yellow) and the mean FA skeleton (blue) derived from the controls + ALS dataset. Note that despite the transformation from FA target space to MNI152 being just affine, the alignment here (with the MNI152) is excellent, as one would hope if the

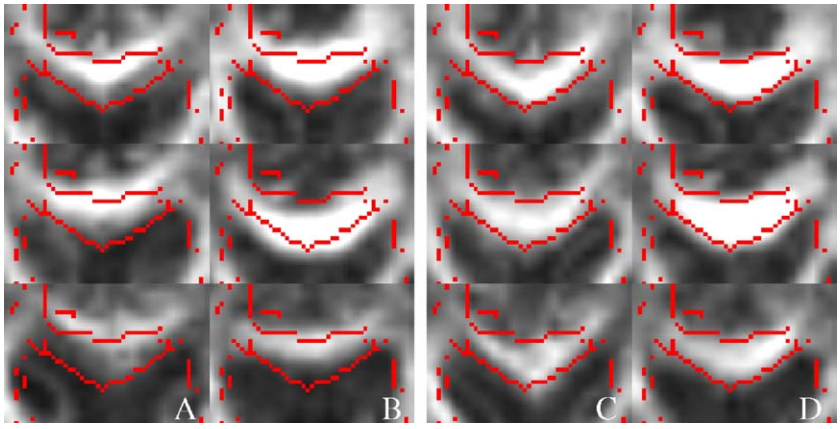

Fig. 6. Example registrations of 3 ALS patients (A, C) and 3 controls (B, D), ROI through the anterior part of the corpus callosum, in axial view. (A, B) Affine-only registration. (C, D) Affine + nonlinear registration. The overlaid red edges are intensity edges from the target image.

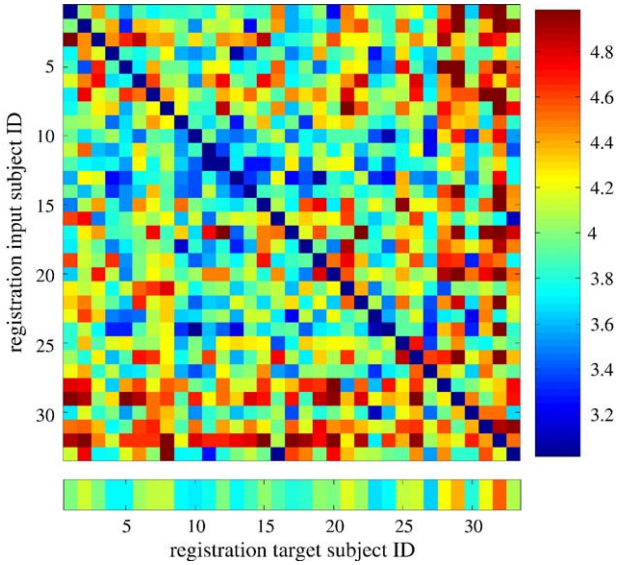

Fig. 7. Example results of summary nonlinear displacement scores (measured in voxels). Each column corresponds to a particular target subject; each row within the column summarises the amount of nonlinear deformation when aligning one of the study's 33 subjects to that target. The bottom row summarises the target subjects; the first 20 subjects are the controls, and the final 13 are ALS patients, with clearly greater structural variability than the control group.

"most typical" subject is generally representative of the wider population.

In Fig. 9, we show the different skeletons created when different subjects are chosen as the target, in order to show the relative stability of the final skeleton against the choice of target subject. The subjects are 20 controls and 13 patients with ALS. In the first analysis, we used all 33 subjects in the alignment target identification; one of the ALS patients was determined to be the most "typical" (subject number 27 in Fig. 7). Next, we used just the 20 controls to find the target subject (number 5) and finally just the 13 ALS patients to find a target subject (number 23). Then, for each of the 3 choices of target subject, we aligned all 33 subjects to the target, formed the mean FA and created the FA skeleton. Fig. 9 shows the 3 skeletons thus formed, shown together and separately, for an example axial and an example coronal slice. All 3 skeletons are thresholded at a mean FA value of 0.3. It is clear that the 3 skeletons are very similar, suggesting that the final skeleton is not sensitive to the set of subjects used in the target space identification or the exact target then selected.

In order to give an idea of the relative number of original and skeletonised white matter voxels, and the effect of thresholding the mean FA skeleton, Fig. 10 shows the skeleton derived from a study comprising 36 controls and 33 schizophrenics, overlaid onto a tissue-type segmentation derived directly from the MNI152 segmentation priors used by SPM and FSL. Green shows voxels with mean FA value in the range 0:0.2; red shows 0.2:0.3, and blue

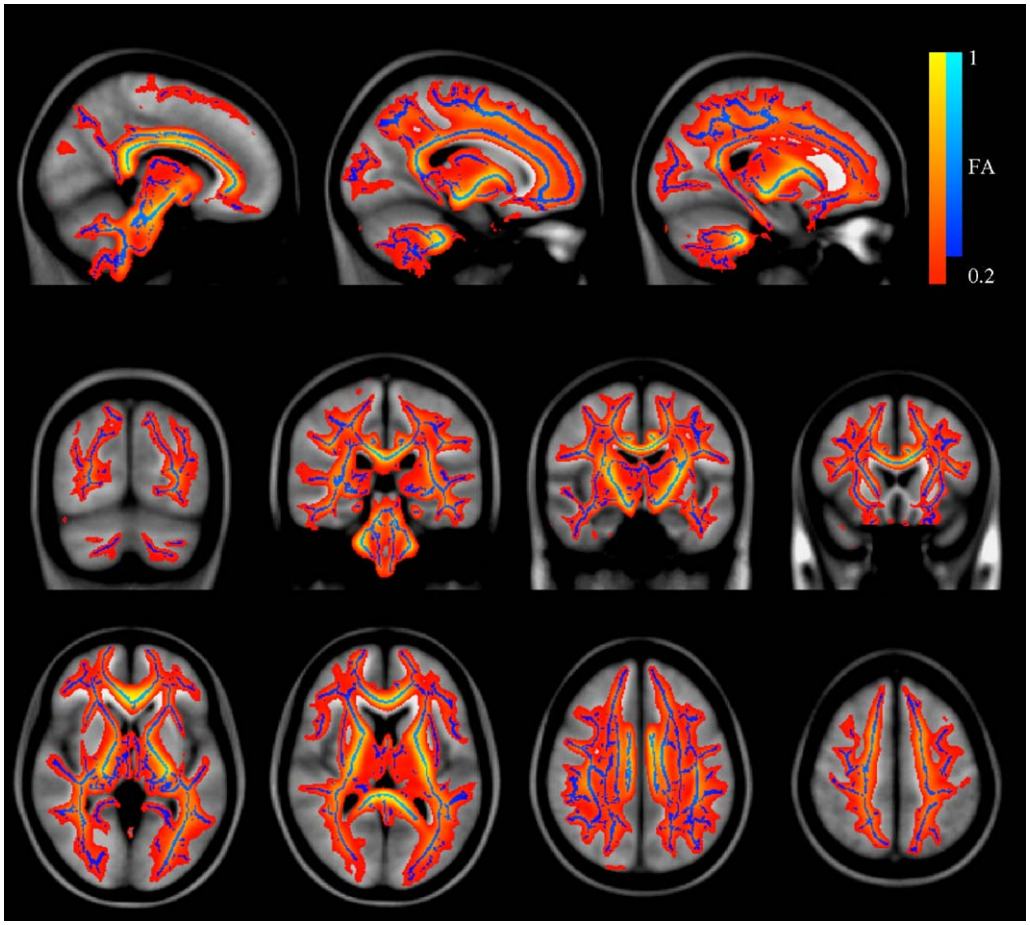

Fig. 8. Example overlay of mean FA map from 20 controls and 13 ALS patients, after each subject has been nonlinearly aligned to the target subject in MNI152 space. The mean FA, shown in red–yellow, is thresholded at 0.2 and overlaid onto the MNI152. The skeleton, shown in blue, is thresholded at 0.3.

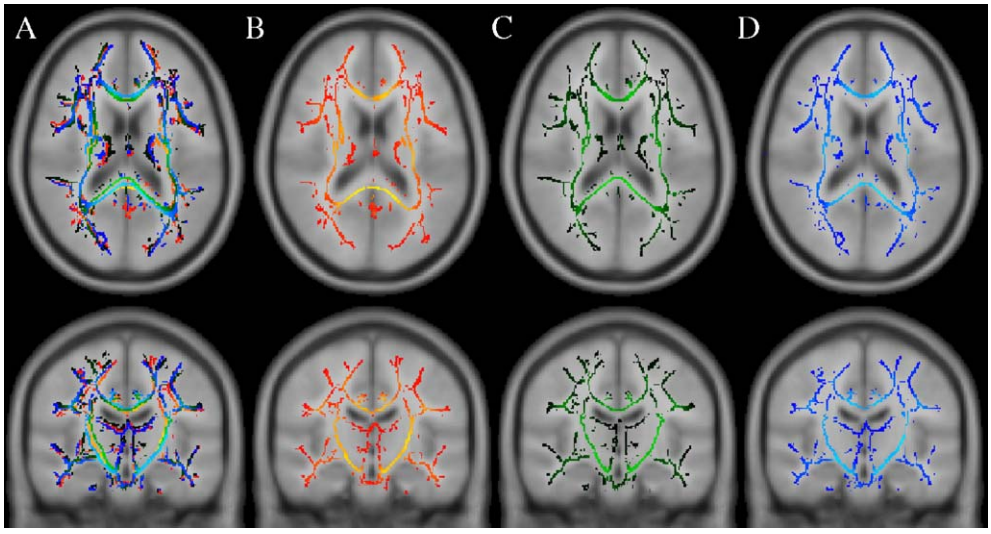

Fig. 9. FA skeletons created using 3 different target subjects for nonlinear registration. (A) All 3 skeletons overlaid. (B) target subject from all 33 subjects. (C) Target subject from just the 20 controls. (D) Target subject from just the 13 ALS patients. All colour maps show FA values from 0.3:1.

shows voxels with FA greater than 0.3 (in the actual schizophrenia results shown later, we used a threshold of 0.2). The number of white matter voxels (which equals the volume in mm^3 at this resolution) in the MNI152 segmentation is 455,154. The total number of skeleton voxels is 289,562; however, the number within the MNI152 white matter mask is 77,374, a sixfold reduction compared with the number in the mask. This reduction reflects the aim of reducing the FA data to being most robustly and informatively represented by just the centres of white matter tracts (though see also the comments in the final discussion relating to the option of also using other measures such as integrated FA or tract width as statistics of interest). With respect to the effect of thresholding, the number of skeleton voxels with FA less than 0.2 is 148,218, of which 146,151 (99%) lie outside the MNI152 white matter mask. Furthermore, of the skeleton voxels inside the MNI152 white matter mask, over 97% have a FA greater than 0.2. These figures show clearly that the general effect of thresholding (at, e.g., 0.2) is to distinguish between areas that are on average grey matter and those that are on average white matter.

Fig. 11 shows the variation in aligned FA images relative to the mean FA skeleton, from a second dataset—15 subjects who stutter and 11 controls. It can clearly be seen that the skeleton lies within or near WM tracts in the great majority of subjects.

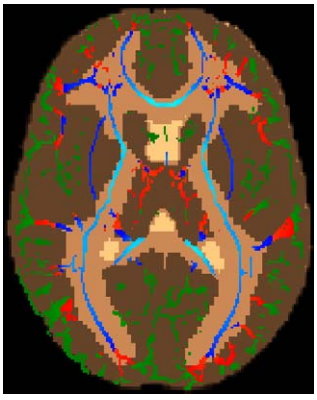

Fig. 10. Mean FA skeleton from 36 controls and 33 schizophrenics, thresholded into three ranges: green = 0:0.2, red = 0.2:0.3, blue = 0.3:1. Underneath is the tissue-type segmentation (into grey, white and CSF) derived from the population-average segmentation priors used by SPM and FSL.

Projecting individual subjects' FA onto the skeleton

Fig. 12 shows the search results in part of an axial slice taken from analysis of 18 normal subjects. For each subject a set of arrows from the skeleton to that subject's (aligned) FA image is shown. It can be seen that where there is slight misalignment of a subject's warped FA image with the skeleton (derived from the mean FA image), the search strategy appears to be correctly taking values from the true centre of the nearest tract. (Note that the search is in 3D so these 2D cross-sectional cuts through the image, and the search vectors do not quite show the whole story.)

In order to show qualitatively an example relationship between tractography output and a mean FA skeleton, we took the reproducibility data (see later) and derived several tracts for a single subject (note: not the same subject as that used as the nonlinear registration target). The tractography was run using FDT (Behrens et al., 2003b; Smith et al., 2004); two masks were defined such that (tract-following) samples were seeded from each mask and accepted only if they passed through the other. After passing through the second mask, the tract following was terminated for clarity of display. Masks were placed by hand in the left and right upper cingulum, optic radiation, cortico-spinal tract and in the genu of the corpus callosum. Fig. 13 shows the 8-subject group mean FA skeleton underneath the tractography output from one of the subjects. On the basis of these images, one would be fairly confident that a perpendicular search from the skeleton voxels will intersect the correct tract appropriately, and it is also clear that the search is necessary to correct the slight misalignment between the tract centre and the skeleton, in several places.

Testing for Gaussianity

As discussed above, it is of interest to test whether projecting data onto the mean FA skeleton improves the Gaussianity of the cross-subject distribution of FA values. In Jones et al. (2005), it was shown that there was a large number of voxels whose cross-

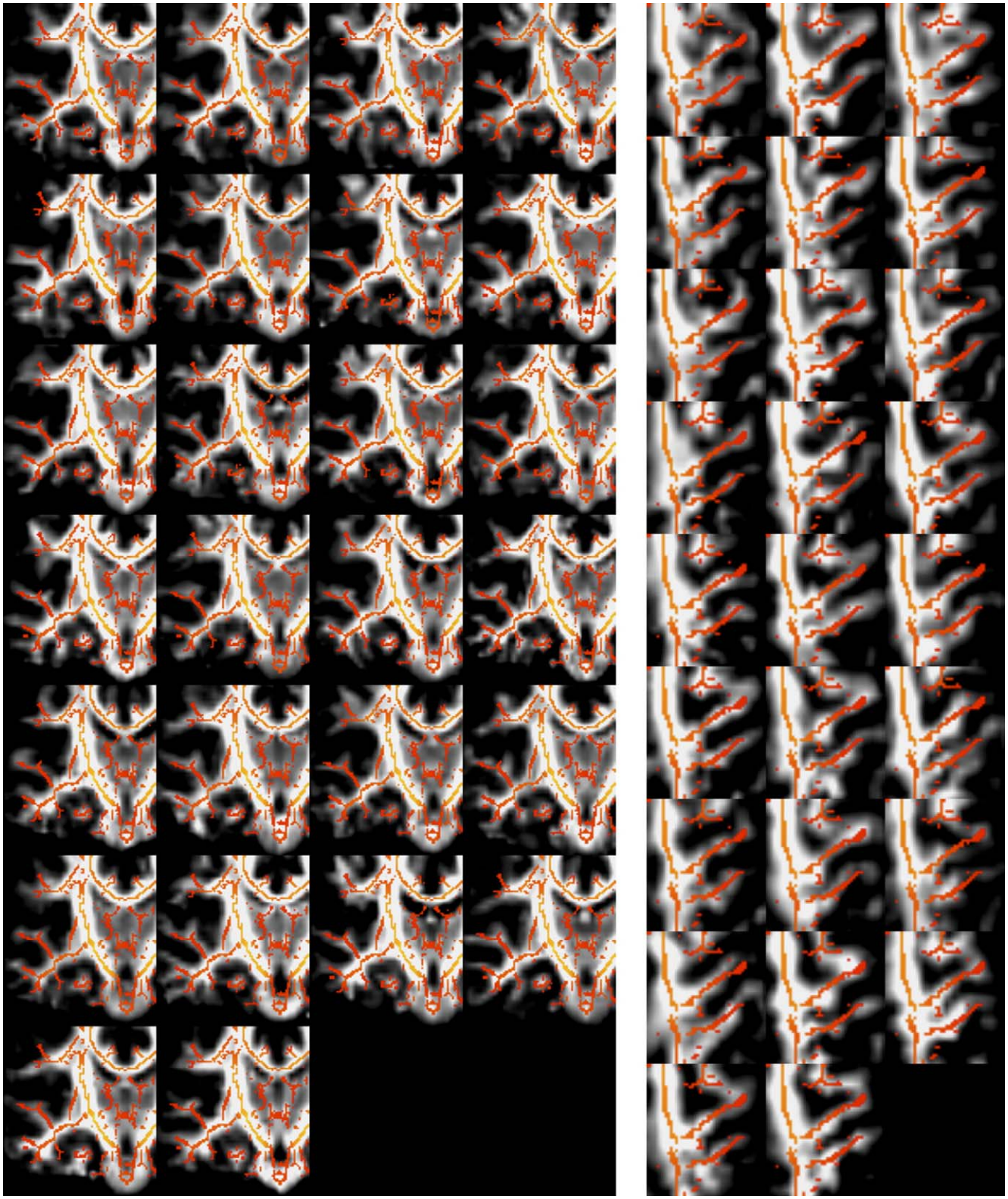

Fig. 11. Individual subject (nonlinearly aligned) FA maps vs. mean FA skeleton in 26 subjects (15 stutterers and 11 controls). Left: coronal ROI; right: axial ROI.

subject distribution was significantly non-Gaussian. We tested two datasets—one comprising 36 controls and the other comprising 33 schizophrenics, using the Lilliefors modification of the Kolmogorov–Smirnov test (Lilliefors, 1967) to find voxels where the cross-subject distribution was significantly non-Gaussian. The test threshold was set at 0.05. Therefore, we expect to find 5% of voxels failing the test by chance; a much higher number of voxels is evidence for non-Gaussianity.

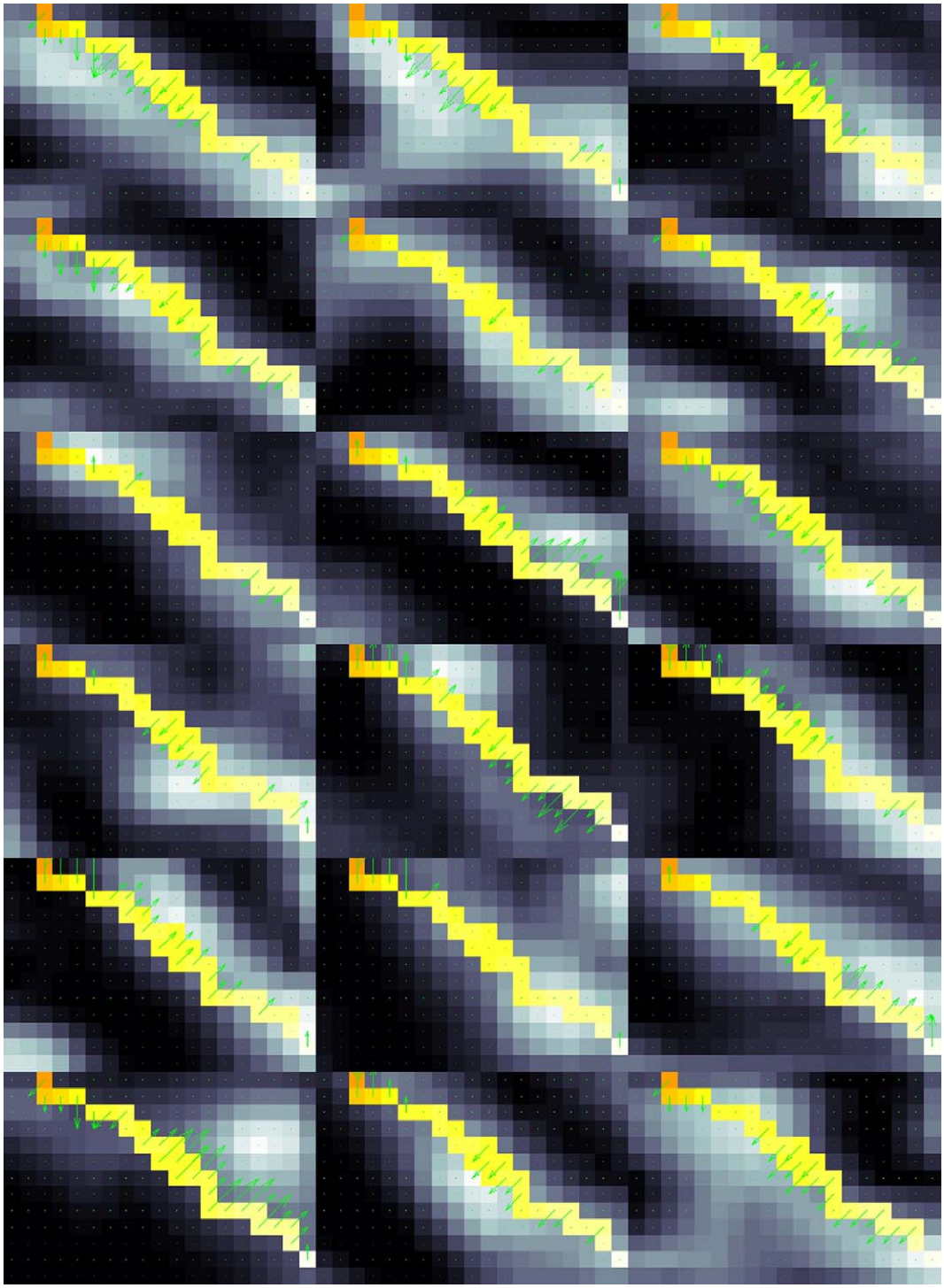

Fig. 12. Axial regions-of-interest showing, for each subject in a group of 18 controls, how each skeleton voxel takes data from the relevant local FA voxel.

We ran the test on each dataset in three ways. Firstly, we tested all voxels after the initial nonlinear registration (and before skeletonisation); this is similar therefore to the VBM-based investigation reported in Jones et al. (2005). Secondly, we masked this aligned data with the mean FA skeleton, and investigated just these voxels—i.e., looking at skeleton voxels, but before projecting the aligned data onto the skeleton. Finally, we tested the skeletonised data after full TBSS preprocessing, i.e., after projection onto the skeleton.

The percentage of voxels found to be non-Gaussian in the controls dataset were (respectively for the three tests): 17.8, 7.0, 6.6. In the schizophrenics dataset, the percentages were 19.2, 8.1, 7.5. Thus, it is clear that with the "VBM-style" analysis, we find a large number of voxels with a non-Gaussian distribution (nearly 4 times more than predicted by chance, in exact agreement with the figure found in Jones et al. (2005) for unsmoothed VBM-preprocessed data). Interestingly, the spatial distribution of these tends to be away from the tract centres, as judged visually, and as

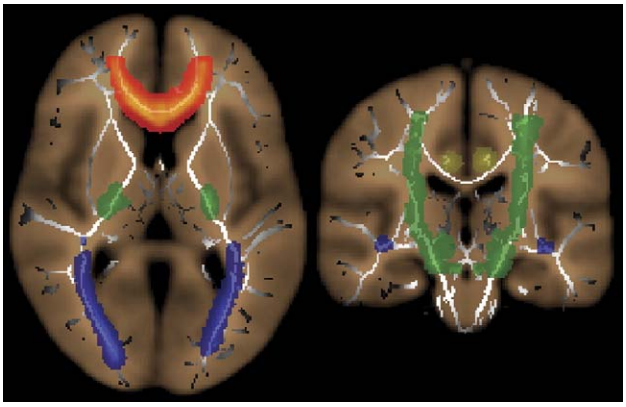

Fig. 13. Example output of probabilistic tractography for several major tracts from a single subject, overlaid on top of the mean FA skeleton derived from 8 normals; the subject used for tractography was not the one used as the reference in the nonlinear registration. In brown is shown the MNI152 average T1 image.

shown by the great reduction in the percentages in the second tests, where the aligned data is only tested at the skeleton voxels. For the fully TBSS-processed data, the test failure rate is reduced still further, to rates not far above the 5% expected by chance.

Repeatability tests

Next, we investigated the repeatability of FA values, both across sessions and across subjects. We analysed data from 8 healthy subjects, each scanned on 3 separate occasions. We estimated % coefficient of variation (CoV: 100 × standard deviation/mean) across sessions or subjects as the measure of repeatability.

We first measured CoV at 7 voxels placed in the centre of various white matter tracts on the mean FA image; the genu of the corpus callosum, left and right optic radiation, left and right pyramidal tract in the cerebral peduncle, and left and right superior cingulum bundle. The exact positioning of the points is described in Heiervang et al. (2006). As well as estimating CoV for the TBSS-preprocessed data at these points, we also found CoV for data before the skeletonisation, after the nonlinear registration stage, which we therefore refer to as being "VBM preprocessed" (though note that no spatial smoothing was applied). Thirdly, we estimated CoV by carefully choosing the relevant voxels of interest by hand on each original FA image separately. Ideally, this hand placing has the advantage of adapting to tract localisation changes across subjects but potentially suffers from subjectivity/user error. In the easiest to define, thickest tracts, hand definition of the voxel in this way should give a close to optimal CoV.

We also obtained global summary statistics (median and mode) across the whole brain for CoV in the TBSS and VBM-preprocessed cases. VBM-preprocessed results are only reported for voxels where the mean FA across all subjects, and all sessions is greater than 0.2, to avoid bias through inclusion of potentially high CoV values in low mean FA voxels. Likewise, the TBSS skeleton was thresholded at the default of 0.2.

Fig. 14 shows the inter-session and inter-subject variability results. Cross-session variability with TBSS preprocessing is generally slightly lower than VBM preprocessing and generally considerably lower than with hand placing. Cross-subject variability with TBSS preprocessing is consistently lower than with VBM preprocessing and lower than hand placing in 4 out of seven positions of interest. The results suggest that TBSS is successful in aligning equivalent structures across sessions/subjects, and that it improves alignment further than pure nonlinear registration has achieved here. With TBSS the inter-session CoV is generally between 3% and 5% (mode 3%), and the inter-subject CoV is generally between 5% and 15% (mode 12%). These figures should prove useful when carrying out power calculations for planned DTI studies.

Example application—schizophrenia

We analysed data from 33 schizophrenics and 36 age-matched controls. After applying the TBSS preprocessing, we first carried out a region-of-interest analysis on mean FA skeleton voxels in the superior cingulum bundle. This was in order to compare our results with those given in Kubicki et al. (2003), where left > right and control > schizophrenic FA differences were reported in the cingulum bundle. Our results were in agreement, namely control > patient ($P = 5.8e{-}3$) and left > right ($P = 8.4e{-}6$).

We then carried out voxelwise statistics using the TBSS-preprocessed data, applying a control-patient unpaired t test. Inference was carried out using cluster-size thresholding, with clusters initially defined by $t > 3$. The null distribution of the cluster-size statistic was built up over 5000 permutations of group membership, with the maximum size (across space) recorded at each permutation. The 95th percentile of this distribution (a cluster size of approximately 150 voxels on the skeleton) was then used

	TBSS % CoV		VBM-preprocessed % CoV		Hand-placed % CoV	
	inter-session	inter-subject	inter-session	inter-subject	inter-session	inter-subject
pointwise - CoV across sessions/subjects at a single voxel						
genu CC	**3.1**	**8.0**	3.3	8.3	9.0	10.3
l optic rad	**4.3**	**11.2**	4.4	12.3	7.7	13.9
r optic rad	**4.8**	**9.4**	4.9	12.2	11.0	13.8
l pyr tract	**3.1**	**4.4**	3.3	5.0	3.2	5.5
r pyr tract	2.5	5.3	**2.4**	6.5	3.6	**3.8**
l cingulum	3.6	6.4	4.6	7.7	**3.2**	**2.8**
r cingulum	4.7	8.8	4.9	11.0	**3.6**	**7.3**
global - CoV across sessions/subjects, computed voxelwise then summarised across space						
mode	**3.2**	**11.8**	3.6	13.5		
median	**5.3**	**15.0**	6.5	21.7		

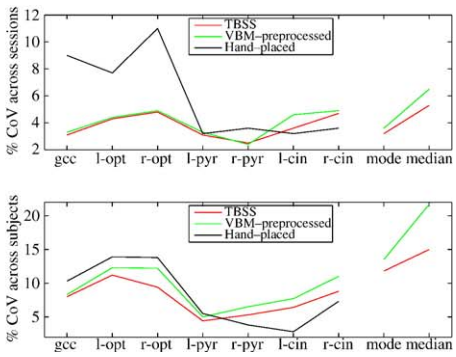

Fig. 14. Inter-session and inter-subject variability results. 8 subjects were scanned 3 times each. Percentage coefficient of variation (CoV) variability results are shown at 7 white matter positions of interest and also using summary statistics for the whole brain. Optimal results for each test are shown in bold.

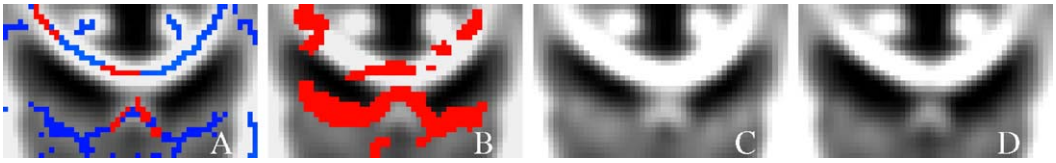

Fig. 15. Coronal views through the controls > schizophrenics group comparison. (A) TBSS analysis showing the FA skeleton in blue and significant group difference in red, in the corpus callosum and fornix. (B) VBM-style analysis, with no spatial smoothing; as well as the corpus callosum and fornix, a group difference is suggested running along the underside of the ventricles. (The 5-mm and 10-mm FWHM smoothing analyses showed the same general pattern, though more diffuse.) (C, D) The mean FA images for the controls and schizophrenics, respectively. It is clear that while the corpus callosum is well aligned between the two groups, the lower edge of the ventricles is not, due to larger ventricles in the patient group. This has given rise to a spurious result in the VBM-style analyses.

as the cluster-size threshold, i.e., the clusters were thresholded at a level of $P < 0.05$, which is fully corrected for multiple comparisons across space (i.e., controlling the familywise error—the chance of one or more false positives *anywhere* on the skeleton).

As well as running TBSS, we also carried out standard VBM-style analysis, using the same nonlinear registration stage. We smoothed the aligned data at a range of spatial extents (0-, 5-, and 10-mm FWHM), carried out the same voxelwise t test as done for the TBSS-preprocessed data, and used the same cluster-size thresholding as described above. The VBM-style analysis was only performed at voxels where the mean FA across subjects (after nonlinear alignment) was greater than 0.15. We considered that any mean FA lower than this is dangerous to consider for a group difference, as such a voxel must be considered to be potentially dominated by grey matter or CSF partial voluming, and any group difference cannot be unambiguously ascribed to change in white matter FA as opposed to a change in relative local amounts of different tissue types.

TBSS found reduced FA in patients in right-superior, medial and anterior corpus callosum, superior and right-inferior fornix and in long association fibres near the junction of the right superior and inferior longitudinal fasciculi. In the majority of these areas, the VBM-style analysis also found a group difference at all 3 spatial smoothing extents, though with much less precision about the exact localisation of group difference. However, in addition, several spurious results were generated by the VBM-style analyses, for example, just below the ventricles, as seen in coronal view in Fig. 15. It is clear from inspecting the mean FA images for the controls and schizophrenics that while the corpus callosum is well aligned between the two groups, the lower edge of the ventricles is not, due to larger ventricles in the patient group. This has given rise to a result which could easily be misinterpreted as a group difference in FA in the VBM-style analyses. TBSS did not show any spurious effect, as it was not sensitive to the between-group shift in this area. For the significant TBSS result in the fornix, we confirmed, through looking at the skeleton projection vectors, that this result was not spurious, i.e., that any inter-subject movement in the fornix was correctly dealt with via the final projection of FA maximum onto the skeleton.

Example application—ALS

We analysed data from 13 ALS patients and 20 controls. After applying the TBSS preprocessing, we carried out two GLM analyses. In the first, using only the patients, we correlated FA with each patient's ALS progression rate, using permutation-based inference on cluster size ($t > 2$, $P < 0.05$ corrected). In the second analysis, we tested where FA was significantly reduced in ALS compared with controls, after regressing out the effect of age (as the two groups were not perfectly age-matched), using permutation-based inference on cluster size ($t > 1$, $P < 0.05$ corrected).

Fig. 16 shows in blue where FA is reduced in ALS compared with controls—the majority of the mean FA skeleton shows reduction, including most of the corpus callosum and pyramidal/corticospinal tracts. Red shows where FA is negatively correlated with ALS progression rate; this is confined to the pyramidal/corticospinal tract, clearly seen in coronal and axial view.

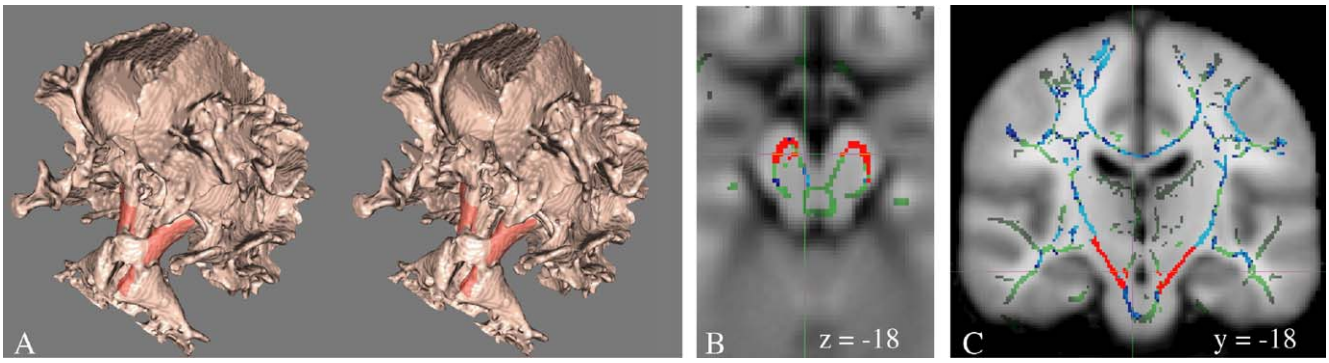

Fig. 16. TBSS results from 13 ALS patients and 20 controls. Red shows where FA correlates negatively with ALS progression rate in the ALS patients. (A) A 3D stereo pair of the mean FA skeleton; to view, cover other parts of the figure, hold approximately 20 cm from the eyes, cross the eyes and slowly focus on the centre fused image. (B, C) Green shows mean FA skeleton, mostly hidden underneath blue and red. Blue (also mostly present "underneath" red voxels) shows where FA is significantly lower in ALS than in controls, after regressing out the effect of age. The background image in panels B, C is the MNI152.

Example application—multiple sclerosis

We analysed data from 15 patients with multiple sclerosis (MS). After applying the TBSS preprocessing, we carried out two GLM analyses. In the first, we correlated FA (voxelwise, across subjects) with each subject's EDSS score (Expanded Disability Status Scale, a common measure of disability), using permutation-based inference on cluster size ($t > 1$, $P < 0.05$ corrected). In the second

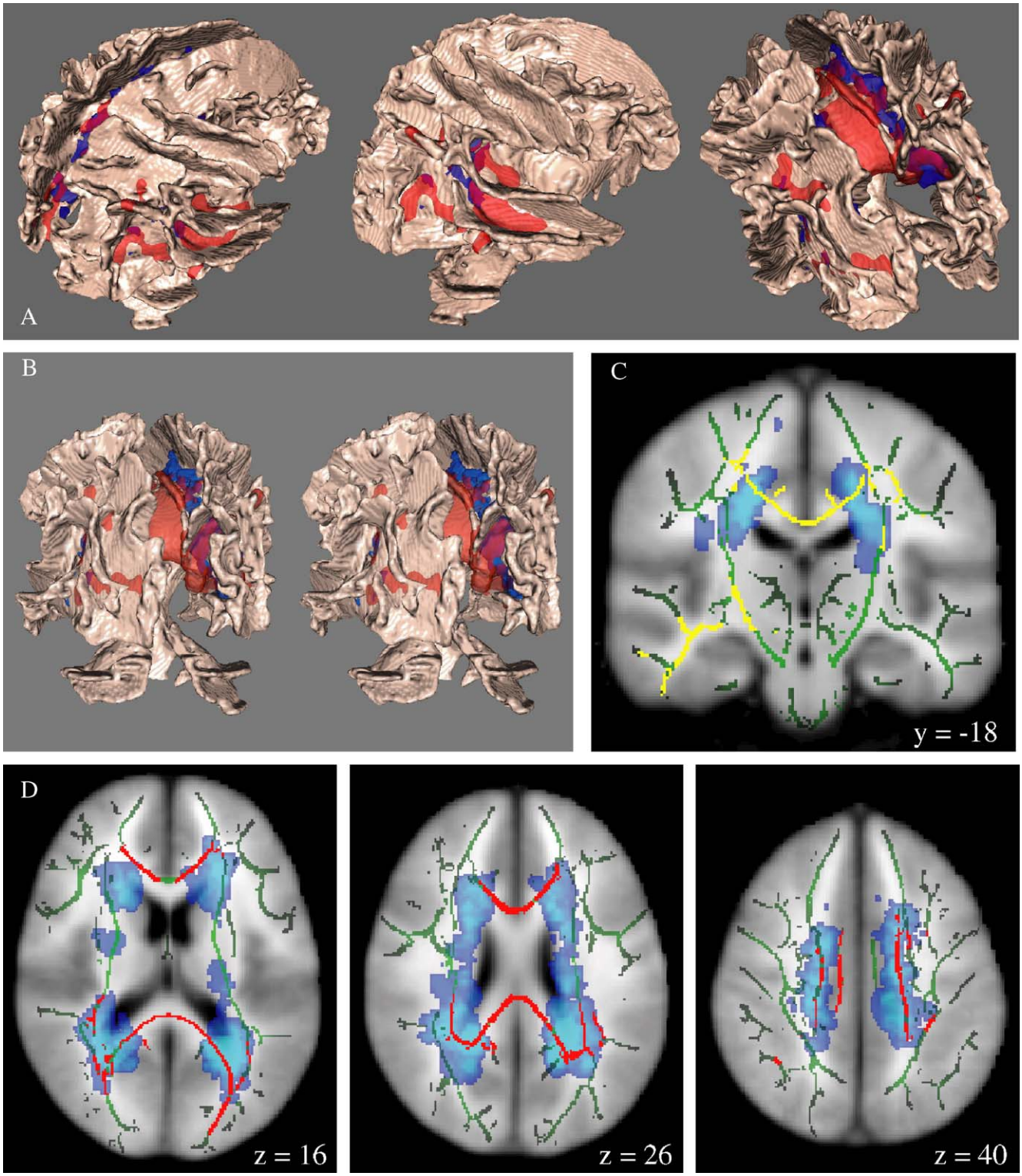

Fig. 17. TBSS results from 15 MS patients. (A, B) 3D surface renderings of the mean FA skeleton. Blue shows the group mean lesion probability distribution, thresholded at 20%. Red shows voxels where FA correlates negatively (across subjects) with subject total lesion volume. Panel B is a 3D stereo pair. (C) Yellow shows where FA correlates negatively with EDSS disability score. (D) Red as above (negative correlation with lesion volume). In panels C and D, green shows the mean FA skeleton, blue shows the group mean lesion distribution, and the background image is the MNI152.

analysis, we correlated FA across subjects with total lesion volume (measured by hand segmentation of T2-weighted images), again with permutation-based inference on cluster size ($t > 2$, $P < 0.05$ corrected).

Fig. 17 shows the mean lesion probability distribution in blue: For each subject, a binary lesion mask is created by hand. All subjects' lesion masks are then transformed into standard space and averaged. The figure shows this mean lesion distribution thresholded at 20% (i.e., at any given blue voxel, 20% of the subjects had a lesion present).

Red voxels on the mean FA skeleton show where FA correlates negatively across subjects with subject total lesion volume. There is strong negative correlation in left superior cingulum and many parts of the corpus callosum, including midline parts of the CC, well away from areas of lesion. This suggests that FA is reduced even in normal appearing white matter as disease progresses.

Yellow voxels show where FA correlates negatively with EDSS disability. Affected areas include superior cingulum, corpus callosum, pyramidal/corticospinal tract, and inferior fronto-occipital/longitudinal fasciculus.

Discussion

In this final section, we discuss some of the limitations of our approach, as well as presenting some potentially interesting areas for future research.

Limitations and dangers

A serious limitation of VBM-style approaches is the need for spatial smoothing, and the problem of arbitrarily choosing the spatial smoothing extent. Another smoothing-related problem lies in the interpretation of cross-subject differences in FA when the white matter is mixed with significant amounts of grey matter—in this case, any estimated change in FA is more likely to be due to a change in the relative amounts of different tissue types than to a change of FA in white matter. See the two foci of detected change in Jones et al. (2005) for an example of this; at least one of these appears to be localised well away from a predominantly white matter area. This problem is greatly exacerbated when applying spatial smoothing, as this increases the mixing of tissue types in any given voxel. However, as one moves away from the larger tracts, this effect will still occur within a voxel even when no smoothing is applied—for example, when tract width is smaller than original voxel size. In this case, it is very difficult to determine whether a reduction in FA is really due to within-tract FA change or a change in tract thickness, and it is important to note that in such cases our approach does not resolve this problem. It is partly for that reason that the mean FA skeleton is thresholded, typically at 0.2, rather than being allowed to fall all the way to zero.

A similar issue is the possible confound of effects such as within-scan head motion. The most obvious effects of increased head motion are increased image blurring and biased FA. This could lead to misinterpretation of apparent subject group differences, if for example a patient group had greater head motion than a control group. Such problems will not in general be resolved through the use of the TBSS approach. One could potentially estimate head motion using image entropy measures and/or motion estimates from the eddy-current/head motion preprocessing and feed this into final statistical analyses as a confound regressor, though this would not be guaranteed to remove all related problems, and could remove the effect of interest.

Another area where careful interpretation is needed is in regions of crossing tracts or tract junctions. As discussed earlier, voxelwise statistics are still difficult to estimate and interpret at tract junctions or crossings. We do not at this point enforce skeleton contiguity at junctions, for practical reasons—a more sophisticated data projection approach would be needed here. In any case, the interpretation of a change in FA at junctions (or areas of crossing tracts) can be complicated; for example, an apparent reduction in FA at junctions can in fact be due to an *increase* in one of the tracts feeding into the junction, if it is a "weaker" tract than others feeding into the junction; see also Jones et al. (2005) and Tuch et al. (2005).

Finally, there is the possibility that pathology could reduce FA so strongly that potential areas of interest may be wrongly excluded from analysis (due for example to the thresholding of the mean FA values on the skeleton). This is in general unlikely, as the effect of most pathologies which are appropriate for cross-subject voxelwise analysis are too subtle (in effect on FA) for this to occur; those pathologies (e.g., gross stroke or large tumours) which would be likely to seriously disrupt tracts (and FA) are unlikely to be suitable for this kind of voxelwise analysis. However, if there was indeed the possibility of the danger of a strong reduction in FA without very large topographical/topological disruption, an appropriate approach would be to use a target FA image for registration, and mean FA skeleton, derived from a relevant control group (ideally a different control group than is part of the study). The final step projecting FA data onto the skeleton would still be expected to be successful in removing residual alignment differences between the different subject categories involved.

Future directions

One obvious area for potential improvement is to use all available diffusion tensor information (rather than just FA), both to drive the preprocessing stages (e.g., alignment), and to feed into the final statistics. For example, Park et al. (2003) show the value in driving nonlinear registration from the full tensor information, and doing this may be worthwhile here, although given the reported improvement in accuracy, the benefits may be modest. It may also be of value to include other imaging (such as T1-weighted structural images) to help drive the alignment. In one study which we analysed with TBSS, the DTI data were of sufficiently bad quality (primarily with respect to signal-to-noise ratio) that we used T1-weighted images instead of FA to drive the nonlinear registration, which did indeed qualitatively improve registration robustness. One could also consider using a white matter segmentation (again output from structural imaging) for the registration, which we would expect to give similar results to using the FA; in this case, we would expect the segmentation-derived images to be lower noise and higher resolution than FA, but possibly containing less rich contrast information.

Furthermore, it would be a natural extension of this work to carry other diffusion measures (mean diffusivity, tensor eigenvalues, principal tract direction, etc.) through the alignment and skeleton projection process and carry out voxelwise statistics on these as well as the FA values (see, for example, Schwartzman et al., 2005). Also, one does not necessarily need to take the maximum FA value when projecting local tract information onto the skeleton; for example, some integration measure of FA within the search space could give an interesting measure of local tract

thickness, though interpretation would need to be made carefully in the light of the previous discussions on partial voluming. Such developments could clearly give a richer set of measures with which to find localised connectivity-related changes across subjects.

It would also be useful and fairly straightforward to define a standard-space skeleton; we have shown in this paper that the "most typical" subject in any given study generally conforms very well to standard space even after purely affine alignment to the MNI152 average image. Hence, a standard-space mean FA image and derived skeleton could simplify TBSS analyses, if one was not concerned about inter-group biases resulting from such a predefined space. A natural extension of this would be to presegment a standard space skeleton into labelled tracts, thus providing the ability to output simple, fully automated reporting of FA statistics within all major tracts as part of the TBSS output.

Finally, there is no reason why one has to carry out the cross-subject statistics separately for each voxel. As with fMRI time series analysis, one could perfectly well feed the entire (sparse) 4D dataset into a multivariate approach such as ICA (independent component analysis (Beckmann and Smith, 2004)), and not only generate added benefit from modelling the spatial aspects of the signal, but potentially find cross-subject modes of variation not predicted in advance.

In this paper, we have presented a new method for estimating localised change in fractional anisotropy, a useful marker for brain connectivity across different subjects. The method attempts to combine the strengths of voxel-based analyses (being able to analyse the whole brain without predefining voxels or tracts of interest) with the strengths of tractography-based analyses (ideally, being confident that the estimates of FA are truly taken from the relevant voxels). We have shown that by projecting FA values onto a subject-mean FA tract skeleton, cross-subject FA becomes more Gaussian and of lower variability; hence analyses become more robust and more sensitive. We have shown example results from applying tract-based spatial statistics to several example datasets. TBSS is freely available as part of FSL (FMRIB Software Library—www.fmrib.ox.ac.uk/fsl).

Acknowledgments

We are grateful to Karla Miller for many discussions regarding the TBSS approach, Andreas Bartsch for valuable advice on this paper, Einar Heiervang for supplying the reproducibility data, and David Flitney and Brian Patenaude for work on the figure generation. We gratefully acknowledge funding from EPSRC, BBSRC, MRC, the Wellcome Trust and the Multiple Sclerosis Society.

References

Ashburner, J., Friston, K., 2000. Voxel-based morphometry—The methods. NeuroImage 11, 805–821.

Ashburner, J., Friston, K., 2001. Why voxel-based morphometry should be used. NeuroImage 14 (6), 1238–1243.

Ashburner, J., Friston, K., 2004. Generative and recognition models for neuroanatomy. NeuroImage 23 (1), 21–24.

Barnea-Goraly, N., Eliez, S., Hedeus, M., Menon, V., White, C., Moseley, M., Reiss, A., 2003. White matter tract alterations in fragile X syndrome: preliminary evidence from diffusion tensor imaging. Am. J. Med. Genet., Part B Neuropsychiatr. Genet. 118B, 81–88.

Basser, P., Matiello, J., Le Bihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson., B 103, 247–254.

Beckmann, C., Smith, S., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imag. 23 (2), 137–152.

Behrens, T., Johansen-Berg, H., Woolrich, M., Smith, S., Wheeler-Kingshott, C., Boulby, P., Barker, G., Sillery, E., Sheehan, K., Ciccarelli, O., Thompson, A., Brady, J., Matthews, P., 2003a. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat. Neurosci. 6 (7), 750–757.

Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003b. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50 (5), 1077–1088.

Bookstein, F., 2001. "Voxel-based morphometry" should not be used with imperfectly registered images. NeuroImage 14 (6), 1454–1462.

Büchel, C., Raedler, T., Sommer, M., Sach, M., Weiller, C., Koch, M., 2004. White matter asymmetry in the human brain: a diffusion tensor MRI study. Cereb. Cortex 14, 945–951.

Cercignani, M., Inglese, M., Pagani, E., Comi, G., Filippi, M., 2001. Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis. Am. J. Neuroradiol. 22, 952–958.

Cercignani, M., Bammer, R., Sormani, M., Fazekas, F., Filippi, M., 2003. Inter-sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. Am. J. Neuroradiol. 24, 638–643.

Conturo, T., Lori, N., Cull, T., Akbudak, E., Snyder, A., Shimony, J., McKinstry, R., Burton, H., Raichle, M., 1999. Tracking neuronal fiber pathways in the living human brain. Proc. Natl. Acad. Sci. U. S. A. 96 (18), 10422–10427.

Davatzikos, C., 2004. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. NeuroImage 23 (1), 17–20.

Ellis, C., Simmons, A., Jones, D., Bland, J., Dawson, J., Horsfield, M., Williams, S., Leigh, P., 1999. Diffusion tensor MRI assesses corticospinal tract damage in ALS. Neurology 53, 1051–1058.

Eriksson, S., Rugg-Gunn, F., Symms, M., Barker, G., Duncan, J., 2001. Diffusion tensor imaging in patients with epilepsy and malformations of cortical development. Brain 124, 617–626.

Gerig, G., Corouge, I., Vachet, C., Krishnan, K., MacFall, J., 2005. Quantitative analysis of diffusion properties of white matter fiber tracts: a validation study. Proc. Int. Soc. of Magnetic Resonance in Medicine, pp. 1337.

Gong, G., Jiang, T., Zhu, C., Zang, Y., Wang, F., Xie, S., Xiao, J., Guo, X., 2005. Asymmetry analysis of cingulum based on scale-invariant parameterization by diffusion tensor imaging. Hum. Brain Mapp. 24, 92–98.

Good, C., Johnsrude, I., Ashburner, J., Henson, R., Friston, K., Frackowiak, R., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14 (1), 21–36.

Guimond, A., Meunier, J., Thirion, J.-P., 2000. Average brain models: a convergence study. Comput. Vis. Image Underst. 77, 192–210.

Heiervang, E., Behrens, T.E.J., Mackay, C.E.M., Robson, M.D., Johansen-Berg, H., 2006. Between session and between subject reproducibility of diffusion MR and tractography measures. ISMRM.

Horsfield, M., 1999. Mapping eddy current induced field for the correction of diffusion weighted echo planar images. Magnetic Resonance Imaging 17, 1335–1345.

Horsfield, M., Jones, D., 2002. Application of diffusion weighted and diffusion tensor MRI to white matter diseases. NMR Biomed. 15, 570–577.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5 (2), 143–156.

Jenkinson, M., Bannister, P., Brady, J., Smith, S., 2002. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17 (2), 825–841.

Jones, D., Horsfield, M., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn. Reson. Med. 42, 515–525.

Jones, D., Griffin, L., Alexander, D., Catani, M., Horsfield, M., Howard, R., Williams, S., 2002. Spatial normalisation and averaging of diffusion tensor MRI data sets. NeuroImage 17, 592–617.

Jones, D., Symms, M., Cercignani, M., Howard, R., 2005. The effect of filter size on VBM analyses of DT-MRI data. NeuroImage 26, 546–554.

Jones, D.K., Catani, M., Pierpaoli, C., Reeves, S.J.C., Shergill, S.S., O'Sullivan, M., Golesworthy, P., McGuire, P., Horsfield, M.A., Simmons, A., Williams, S.C.R., Howard, R.J., 2006. Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. Hum. Brain Mapp. 27 (3), 230–238.

Kochunov, P., Lancaster, J., Hardies, J., Thompson, P., Woods, R., Cody, J., Hale, D., Laird, A., Fox, P., 2005. Mapping structural differences of the corpus callosum in individuals with 18q deletions using targetless regional spatial normalization. Hum. Brain Mapp. 24, 325–331.

Kubicki, M., Westin, C.-F., Nestor, P., Wible, C., Frumin, M., Maier, S., Kikinis, R., Jolesz, F., McCarley, R., Shenton, M., 2003. Cingulate fasciculus integrity disruption in schizophrenia: a magnetic resonance diffusion tensor imaging study. Biol. Psychiatry 54, 1171–1180.

Le Bihan, D., 2003. Looking into the functional architecture of the brain with diffusion MRI. Nat. Rev., Neurosci. 4 (6), 469–480.

Lilliefors, H., 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc. 62, 399–402.

Lim, K., Helpern, J., 2002. Neuropsychiatric applications of DTI—A review. NMR Biomed. 15, 587–593.

Moseley, M., 2002. Diffusion tensor imaging and aging—A review. NMR Biomed. 15, 553–560.

Moseley, M., Cohen, Y., Kucharaczyj, M., Mintorovitch, J., Asgari, H., Wendland, M., Tsuruda, J., Norman, D., 1990. Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. Radiology 176, 439–445.

Neil, J., Miller, P., Mukherjee, P., Hüppi, P., 2002. Diffusion tensor imaging of normal and injured developing human brain—A technical review. NMR Biomed. 15, 543–552.

Nichols, T.E., Holmes, A.P., 2001. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Pagani, E., Filippi, M., Rocca, M., Horsfield, M., 2005. A method for obtaining tract-specific diffusion tensor MRI measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis. NeuroImage 26 (1), 258–265.

Pantazis, D., Nichols, T., Baillet, S., Leahy, R., 2005. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. NeuroImage 25 (2), 383–394.

Park, H.-J., Kubicki, M., Shenton, M., Guimond, A., McCarley, R., Maier, S., Kikinis, R., Jolesz, F., Westin, C.-F., 2003. Spatial normalization of diffusion tensor MRI using multiple channels. NeuroImage 20, 1995–2009.

Park, H.-J., Westin, C.-F., Kubicki, M., Maier, S., Niznikiewicz, M., Baer, A., Frumin, M., Kikinis, R., Jolesz, F., McCarley, R., Shenton, M., 2004. White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. NeuroImage 23, 213–223.

Pierpaoli, P., Basser, P., 1996. Toward a quantitive assessment of diffusion anisotropy. Magn. Reson. Med. 36, 893–906.

Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imag. 18 (8), 712–721.

Rugg-Gunn, F., Eriksson, S., Symms, M., Barker, G., Duncan, J., 2001. Diffusion tensor imaging of cryptogenic and acquired partial epilepsies. Brain 124, 627–636.

Schwartzman, A., Dougherty, R., Taylor, J., 2005. Cross-subject comparison of principal diffusion direction maps. Magn. Reson. Med. 53, 1423–1431.

Simon, T., Ding, L., Bish, J., McDonald-McGinn, D., Zackai, E., Gee, J., 2005. Volumetric, connective, and morphologic changes in the brains of children with chromosome 22q11.2 deletion syndrome: an integrative study. NeuroImage 25, 169–180.

Smith, S., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.

Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., De Luca, M., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J., Matthews, P., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (S1), 208–219.

Sommer, M., Koch, M., Paulus, W., Weiller, C., Büchel, C., 2002. Disconnection of speech-relevant brain areas in persistent developmental stuttering. The Lancet 360, 380–383.

Tuch, D., Salat, D., Wisco, J., Zaleta, A., Hevelone, N., Rosas, H., 2005. Choice reaction time performance correlates with diffusion anisotropy in white matter pathways supporting visuospatial attention. Proc. Natl. Acad. Sci. U. S. A. 102, 12212–12217.

Vangberg, T., Kristoffersen, A., Tuch, D., Dale, A., Skranes, J., Brubakk, A.-M., Larsson, H., Harald-seth, O., 2005. White matter diffusion anisotropy in adolescents born prematurely. Proc. Int. Soc. of Magnetic Resonance in Medicine. , pp. 296.

Watkins, K., Vargha-Khadem, F., Ashburner, J., Passingham, R., Friston, K., Connelly, A., Frackowiak, R., Mishkin, M., Gadian, D., 2002. MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain 125, 465–478.

Worsley, K., Evans, A., Marrett, S., Neelin, P., 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metab. 12, 900–918.