

Twenty-five Pitfalls in the Analysis of Diffusion MRI Data[†]

Derek K. Jones^{a*} and Mara Cercignani^b

Obtaining reliable data and drawing meaningful and robust inferences from diffusion MRI can be challenging and is subject to many pitfalls. The process of quantifying diffusion indices and eventually comparing them between groups of subjects and/or correlating them with other parameters starts at the acquisition of the raw data, followed by a long pipeline of image processing steps. Each one of these steps is susceptible to sources of bias, which may not only limit the accuracy and precision, but can lead to substantial errors. This article provides a detailed review of the steps along the analysis pipeline and their associated pitfalls. These are grouped into (1) pre-processing of data; (2) estimation of the tensor; (3) derivation of voxelwise quantitative parameters; (4) strategies for extracting quantitative parameters; and finally (5) intra-subject and inter-subject comparison, including region of interest, histogram, tract-specific and voxel-based analyses. The article covers important aspects of diffusion MRI analysis, such as motion correction, susceptibility and eddy current distortion correction, model fitting, region of interest placement, histogram and voxel-based analysis. We have assembled 25 pitfalls (several previously unreported) into a single article, which should serve as a useful reference for those embarking on new diffusion MRI-based studies, and as a check for those who may already be running studies but may have overlooked some important confounds. While some of these problems are well known to diffusion experts, they might not be to other researchers wishing to undertake a clinical study based on diffusion MRI. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: diffusion tensor MRI; registration; distortion; spatial normalization; analysis; statistics; artefacts; voxel-based analysis

INTRODUCTION

Diffusion-weighted (DW) (1,2), and diffusion tensor (DT) MRI (3,4) are amongst the most popular imaging techniques for assessing brain tissue integrity, and white matter in particular. To date, diffusion MRI is the only non-invasive method for probing tissue microstructure *in vivo* and offers the possibility of exploring the connectivity of distinct anatomical networks within the brain (5). With the increasing realisation that to fully understand the function of the brain, one has to also understand its structure, it is no surprise that the applications of diffusion MRI have exploded in recent years.

Obtaining reliable data and drawing meaningful and robust inferences from them, with diffusion MRI, however, can be challenging. The process of quantifying diffusion indices, and eventually comparing them between groups of subjects and/or correlating them with other parameters starts at the acquisition of the raw data, followed by a long pipeline of image processing steps, containing many potential pitfalls. Each one of these steps is susceptible to sources of bias, which may not only limit the accuracy and precision, but can lead to substantial errors.

Here we review in detail several of the steps along the analysis pipeline, highlighting potential pitfalls on the way. These are grouped into (1) pre-processing of data; (2) estimation of the tensor; (3) derivation of voxelwise quantitative parameters; (4) strategies for extracting quantitative parameters; and finally (5) intra-subject and inter-subject comparison.

Note that we will not be covering fibre tracking/connectivity assessments in depth, as this field contains many pitfalls and confounds and their coverage would double the length of this article. An article focused on the pitfalls of fibre tracking is planned in the future.

Here we have assembled 25 pitfalls (several previously unreported) into a single article, which should serve as a useful reference for those embarking on new diffusion MRI-based studies, and as a check for those who may already be running studies but may have overlooked some important confounds. In selecting 25 pitfalls, we are by no means suggesting that we have assembled an exhaustive list. The reader may have already succumbed to others, and many more will be sure to emerge. The aim of this article, however, is not to suggest that diffusion MRI is

* Correspondence to: D. K. Jones, CUBRIC, School of Psychology, Cardiff University, Park Place, Cardiff, CF10 3AT, UK.
E-mail: jonesd27@cf.ac.uk

a D. K. Jones
CUBRIC, Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff, UK

b M. Cercignani
Neuroimaging Laboratory, Santa Lucia Foundation, IRCCS, Rome, Italy

† This article is published in NMR in Biomedicine as a special issue on Progress in Diffusion-Weighted Imaging: Concepts, Techniques, and Applications to the Central Nervous System, edited by Jens H. Jensen and Joseph A. Helpern, Center for Biomedical Imaging, Department of Radiology, NYU School of Medicine, New York, NY, USA.

Abbreviations used: ADC, Apparent diffusion coefficient; b, b-value; B, Bmatrix; B_0 , static magnetic field; DT, Diffusion Tensor; DW, Diffusion-weighted; EC, Eddy current; EPI, Echo-planar imaging; FA, Fractional anisotropy; FLAIR, fluid attenuated inversion recovery; FWHM, full width at half maximum; HARDI, high angular resolution diffusion imaging; MSE, mean squared error; NLLS, nonlinear least squares; OLS, ordinary least squares; PET, positron emission tomography; RA, relative anisotropy; RF, radio frequency; ROI, region of interest; SNR, signal-to-noise ratio; TBSS, tract based spatial statistics; TE, echo time; VB, voxel based; WLS, weighted least squares.

unreliable or should not be used. On the contrary, we hope that the variety of topics that we have selected for this review will motivate the reader to consider their analysis pipelines more closely, which will ultimately lead to increased robustness of diffusion studies within a laboratory, and therefore increased reliability of cross-centre comparisons of diffusion-MRI results.

PRE-PROCESSING OF DT-MRI DATA

With 'pre-processing' we refer to any manipulation of the raw data performed prior to the fitting of a model. Discussing data acquisition is beyond the scope of this article; however, it should be mentioned that the type of image artefacts which are likely to occur in DT MRI datasets is largely determined by the type of pulse sequence and acquisition strategy used. Echo-planar imaging (EPI) is by far the most common readout strategy employed in DT MRI, although several alternative methods exist, as it enables robust DW images to be acquired without significant motion artefacts. However, as multiple images are acquired for estimation of the tensor, involuntary motion between images can pose a problem. Moreover, the rapid readout of k-space in EPI leads to a low bandwidth in the phase-encode direction, which also makes the images extremely sensitive to off-resonance, susceptibility and eddy current effects. The pre-processing part of the pipeline should aim to minimize the effect of each of these confounds. The advances in technology in the latest years, with the advent of multi-channel arrays, the development of acquisition methods based on parallel imaging (6–8), and of sequences which allow a better cancellation of eddy currents (9) have largely reduced the effects of off-resonance artifacts on DT MRI. Nevertheless, it is important to be aware of these problems, and to understand the shortcomings of some widely adopted correction methods.

Susceptibility effects and correction strategies

When a sample is comprised of different components (e.g. tissue, air, bone), each characterised by a different magnetic susceptibility, the magnetic field B_0 is locally altered and thus shows discontinuities at tissue interfaces. In the brain, the most obvious example can be observed in the temporal lobes, because of the vicinity of the paranasal sinuses. This discontinuity in susceptibility leads to two macroscopic effects in EPI: a shift of the maximum signal away from the theoretical echo time, which causes signal loss, and geometric distortions. Signal drop out typically occurs in gradient echo images, and thus is less of an issue for DW spin echo EPI. With respect to geometric distortion, however, field inhomogeneity behaves like a 'background gradient' that affects the phase of the measured signal. It was shown (10) that k-space sampled by an EPI readout in the

presence of field inhomogeneity ΔB is the Fourier transform of an image where voxels are displaced by a quantity proportional to ΔB , and modulated by a term dependent (to a first approximation) on the first derivative of ΔB with respect to the phase encoding direction. This displacement clearly alters anatomical fidelity, and makes co-registration between DT MRI parametric maps and other types of images difficult to achieve.

Many strategies to deal with this problem have been proposed (e.g. (11–15)): in particular the effects of susceptibility can be mitigated by combining EPI with parallel imaging to reduce the duration of the echo train (15) and increasing the bandwidth. This, however, will only reduce and not remove the problem. The most commonly adopted solution consists of measuring the B_0 field and applying a retrospective correction to the distorted data (10,12). The direct measurement of the B_0 field can be achieved through an additional standard acquisition, typically an asymmetric spin-echo or a gradient echo, where the phase evolution is proportional to $\Omega = (\omega + \gamma\Delta B) TE$ (where ω is the Larmor frequency at the undistorted B_0 field, γ is the gyromagnetic ratio, and TE is the echo time). An estimate of ΔB can be obtained by acquiring two scans with different echo times, and then taking the difference of their phases. Such a field map allows the local shift to be calculated voxel-by-voxel, and geometric distortion to be corrected. This procedure is also known as 'unwarping'.

PITFALL 1: MANY-TO-ONE MAPPING DUE TO SUSCEPTIBILITY-INDUCED DISTORTIONS

As susceptibility-induced distortions are non-linear it is possible that the signal intensity from neighbouring voxels collapses into a single voxel, in other words, the mapping between the original and the distorted space is 'many-to-one'. In this case unwarping becomes an ill-posed problem. Although theoretically it is possible to use the field map to estimate the spatial derivative of ΔB , and therefore to correct image intensity as well as distortions, in practice such estimates tend to be inaccurate where information is most needed, due to the rapid susceptibility variation across the object. Retrieval of the original signal intensity distribution is thus problematic.

An example consequence of this pitfall is shown in Figure 1: images of FA for the same subject and the same slice location are displayed, obtained before and after unwarping through a field map correction. A T_2 -weighted anatomical image of the same slice is also shown for comparison. After the unwarping, two structures appear (indicated by the arrows) that are perfectly straight and have completely uniform anisotropy along them. These structures are not consistent with any known anatomy. A more likely explanation for their appearance is that the signals from several voxels collapsed into a single voxel during the acquisition, due to field inhomogeneity, and these signals were

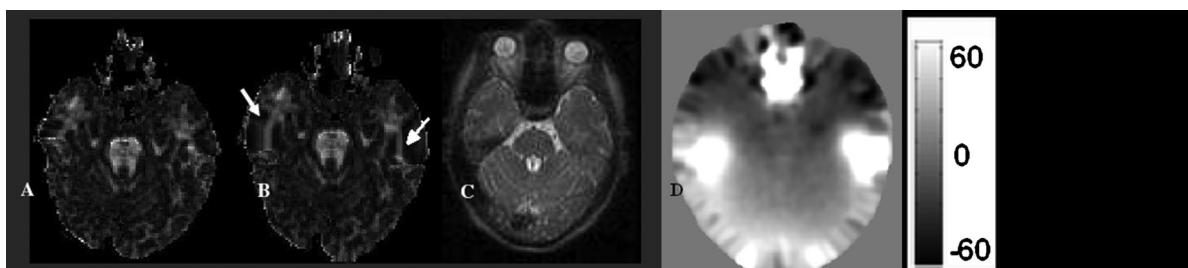


Figure 1. FA maps computed from data before (A) and after (B) unwarping using the standard field mapping approach. A T_2 -weighted anatomical image (C) is also shown for comparison. Note the appearance of the perfectly straight 'fiber pathways' in the unwarped image. The anisotropy is uniform along these segments. These are most likely an artifact due to the many-to-one mapping. The corresponding B_0 map in Hertz is shown in (D).

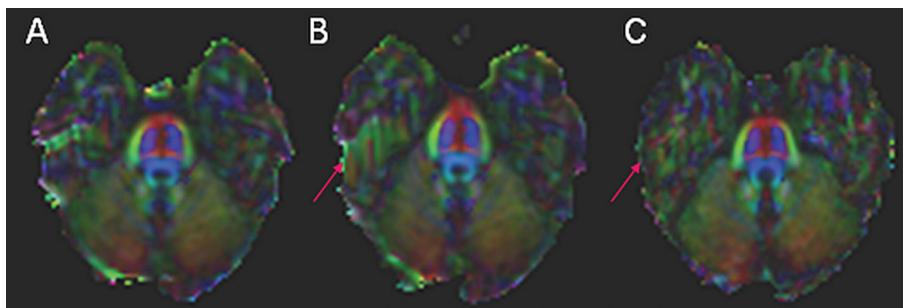


Figure 2. Colour-coded direction maps in the presence of susceptibility distortion: before correction (A), after correction using the standard field-mapping method (B) and after correction using the reversed gradient method (C). The directional information is better preserved when using the reversed gradient method, particularly in the area pointed by the arrows. Note that also eddy current distortions are better compensated in C

incorrectly ‘demodulated’. As a result the same signal is assigned to each of the original voxels—resulting in identical tensors being computed in each.

An alternative approach to correcting for susceptibility effects consists of the acquisition of two data sets that are otherwise identical but with opposite k-space traversal direction [reversed gradient method (11)], and therefore opposite distortions, which can later be combined to form a distortion-corrected image (11–14) without loss of information, at the price of collecting twice as many data as those required by the experiment. The orientational information is better preserved when using this method as opposed to field mapping (see Fig. 2). Moreover, the additional information provided by the two images obtained with phase encoding gradients of opposite polarity can be exploited to simultaneously correct for susceptibility, eddy-current and motion, if these effects are combined in a single model (16).

PITFALL 2: FIBER TRACKING IN THE PRESENCE OF SUSCEPTIBILITY-INDUCED DISTORTIONS

Another pitfall concerns fibre tracking in the presence of susceptibility artefacts where nonlinear distortions can affect the outcome of tractography algorithms (17,18). Examples of results of tractography performed before and after correction have been presented by Andersson *et al.* (17) and Lee *et al.* (18). After correction, the reconstructed tracts appear more symmetric than before and consistent with known anatomy. In some cases, they might even end in different sulci than before correction. Therefore, compensating for susceptibility effects before fibre tracking is a highly desirable, albeit not yet widely adopted, strategy. Field mapping, however, (for the reasons mentioned above) might not be the most suitable way, and a careful inspection of results is definitely recommended.

Correcting for eddy current induced distortions

Another issue related to the low bandwidth in the phase-encode direction, particularly using echo-planar readouts, is that of eddy currents. When a magnetic field is time-varying such as ramping up a diffusion-encoding gradient, electric currents (eddy currents) will be generated in nearby conductors, generating local magnetic field gradients that will either add to or subtract from the gradients that are used for spatial encoding. In most imaging acquisitions, eddy currents are not a major problem. Spatial encoding gradients are normally applied for short periods—such that the rising and falling edges of the gradient are close together in time, and thus there is a form of self-compensation. Due to limited gradient amplitude on clinical systems, however, achieving b-values required for robust diffusion tensor imaging (e.g. of the order of 1000 s mm^{-2}) or for

HARDI-type acquisitions (of the order of $2000\text{--}3000 \text{ s mm}^{-2}$), requires that the gradients are on for much longer than usual. Thus, the rising and falling parts of the gradient waveform are sufficiently temporally separated that the eddy currents are no longer self-compensated. The low bandwidth in the phase-encode direction means that a perturbation of the local magnetic field is going to have a much larger effect on image distortions than normally seen with other imaging readout methods.

If the time-constant of the eddy current is long (by which we mean that it effectively remains constant during the course of the EPI readout train), then the distortions that are produced can be easily predicted. First, all distortions will be along the phase-encode direction, where the bandwidth is low. (The bandwidth in the read-direction is effectively infinite.) Second, a residual eddy current along the slice-direction will mean that every voxel in the slice is subject to the same phase shift and so there will be a uniform shift of the image slice along the phase-encode direction. If there is a residual eddy current along the read-direction, then the artefactual phase-offset due to the eddy current will be dependent on the position along the read-axis. This will lead to a shear of the image along the phase encode direction. Finally, if there is a residual gradient along the phase encode direction, this will lead to a translation of the voxel that is dependent on the position along the phase encode direction. In other words, there will be a stretch or a compression of the image. The eddy current distortions will vary with the diffusion encoding applied, and thus there will be mis-registration between successive images used to estimate the tensor. At the edge of the brain, this will lead to a diffusion-weighted signal that varies rapidly with the direction of the encoding gradient—and thus will appear highly anisotropic. Thus, the characteristic ‘tell’ of eddy currents is a rim of high anisotropy along the phase encode direction (Fig. 3). The authors note that they have overheard conversations at conferences where it has been discussed that as long as one performs analyses away from these rim artefacts, then eddy currents will not be a concern. In reality, every single voxel in the data set is subject to eddy current mis-registration and so this assertion is incorrect.

Prevention is always better than cure, and while it is best to try and minimize eddy currents at the acquisition stage (e.g. 9), residual distortions will almost always remain. Provided the time constant of the eddy currents is long (see above), then it is possible to use simple image processing methods [slice-by-slice global affine transformations (e.g. (19,20))] to correct for these forms of eddy current induced distortion. For example, to correct for residual slice-axis eddy currents each diffusion-weighted image can be trivially translated to the non-DW image. Similarly, the shear part of an affine transformation can correct for residual

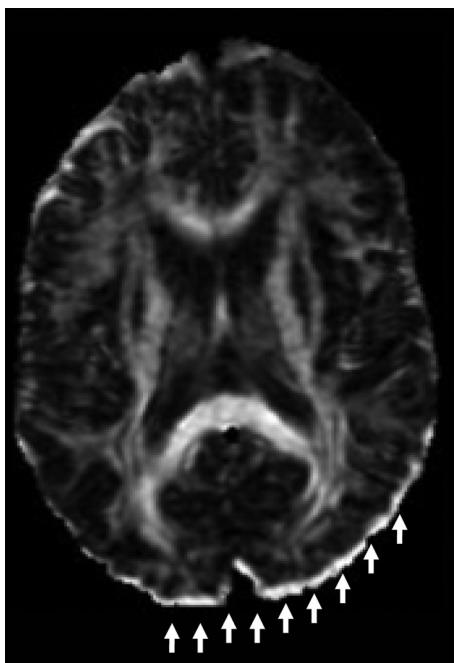


Figure 3. Typical eddy-current induced rim of high anisotropy along the direction of the phase-encoding gradient. Note that the artifact is most prominent at the rim, since these voxels will sometimes be occupied by tissue, and sometimes by air, leading to large variations in signal intensity with encoding direction. However, it should be noted that all voxels in the data set are corrupted by eddy currents, if not corrected.

read-axis eddy currents. Finally, the distortions arising from phase-encode axis eddy currents, can be corrected by applying a stretch/compression along the phase encode axis. For this reason, it has become standard practice to use off-the-shelf registration packages to correct for eddy current distortions by simply applying a global affine transformation to warp each diffusion-weighted volume to the non-diffusion-weighted volumes. However, in using such packages a couple of pitfalls need to be avoided.

PITFALL 3: USING A GLOBAL AFFINE TRANSFORMATION FOR CORRECTION OF EDDY CURRENT DISTORTIONS

The first pitfall in eddy current correction comes from blindly using a single affine transformation for the whole head. Eddy current induced distortions will depend on the slice position and the order in which the slices are played out (see ref. 21 for examples of slice-to-slice variation). Consequently, failing to correct slice-by-slice will mean that only the average transformation is derived—and each and every image will therefore be sub-optimally corrected. While some approaches do correct slice by slice (e.g. (19–20)), the most widely used tools in the field do not, and use a whole brain global affine transformation. It is our impression that this decision is not always an informed and conscious decision made by end-users of packages, but is more driven by availability and ease of use of certain software packages and a lack of understanding of their shortcomings and alternatives.

PITFALL 4: INTENSITY MODULATION DUE TO RESIDUAL PHASE-ENCODING EDDY CURRENTS

The second pitfall in EC correction, also largely ignored in the field, arises when correcting for residual phase-encode eddy currents. As described above, such eddy currents will lead to a

stretch or compression of the image along the phase-encode direction. The amount of stretch/compression will be linearly proportional to the amplitude of the encoding gradient that is causing the eddy current [a fact exploited in one of the original methods for post-processing EC correction (19)]. Correcting for the shape of the voxel is, however, only half of the issue. Whenever there is a change in the volume of the voxel, there will be a change in the signal intensity. For example, if the voxel doubles in volume, then the signal intensity will drop by half. Consequently, when reversing the stretch/compression of the voxel, one should also modulate the signal intensity according to volumetric change (21). If this is not done, the change in signal intensity that results will be ‘interpreted’ by the tensor regression as a change in the underlying diffusivity, leading to a bias. If the matrix deformation that maps the distorted image to the undistorted image is the Jacobian matrix, J , then the volumetric change is given by the determinant of the Jacobian, $\det(J)$. Until now, there has been no quantitative assessment of the consequences of neglecting to perform this signal modulation step on parameters derived from diffusion MRI. For the purposes of this review, we have therefore simulated the consequences of neglecting this step.

Taking data from Horsfield (22), we see that the volumetric change of the voxel can be as great as 8%. Thus, we first simulated two diffusion tensors—with a fixed trace ($2.1 \times 10^{-3} \text{ mm}^2 \text{s}^{-1}$) and two different levels of fractional anisotropy ($FA = 0.3; 0.7$) and a fixed orientation in spherical co-ordinates (theta, phi). Taking an optimal set of 30 direction sampling vectors—‘Jones30’ (23) and assuming a b-value of 1000 s mm^{-2} , we computed the noise free diffusion-weighted intensities and then modulated the signal intensity according to the amount of x-gradient applied [assuming maximal stretch of the voxel of 8% (22) when full power is applied to the gradient]. We then recomputed the tensor with these modulated intensities and compared the estimated trace, anisotropy and principal eigenvector in the original and ‘corrupt’ data sets. This was then done for a range of orientations (theta, phi). Typical results, showing the error in eigenvector orientation, are presented in Figure 4.

There are two main observations. The first is that the error is dependent on the anisotropy. In general, tensors with lower anisotropy will be more susceptible to this artefact than higher anisotropy. This is easily understood by considering two extreme cases. The first is a tensor where there is only one non-zero eigenvalue, and $FA = 1$. A stretch of the image when the gradient is applied perpendicular to the long axis of this tensor will lead to additional attenuation of the signal—but the effect will be minor, and the principal eigenvector will continue to point along the long axis. At the other end of the spectrum, the second case is a tensor where the anisotropy is close to zero, but the principal eigenvector points along the y-axis. If there is stretch of the image when a gradient is applied along the x-axis, the additional attenuation of the image could make the principal eigenvector now switch direction completely to point along the x-axis. The second is that the bias in the estimated eigenvector depends on the fibre orientation—with the result that the bias will be heterogeneously distributed throughout the brain.

Correction for subject motion

There is a growing trend in diffusion MR acquisitions to increase the number of measurements taken at each spatial location. Most

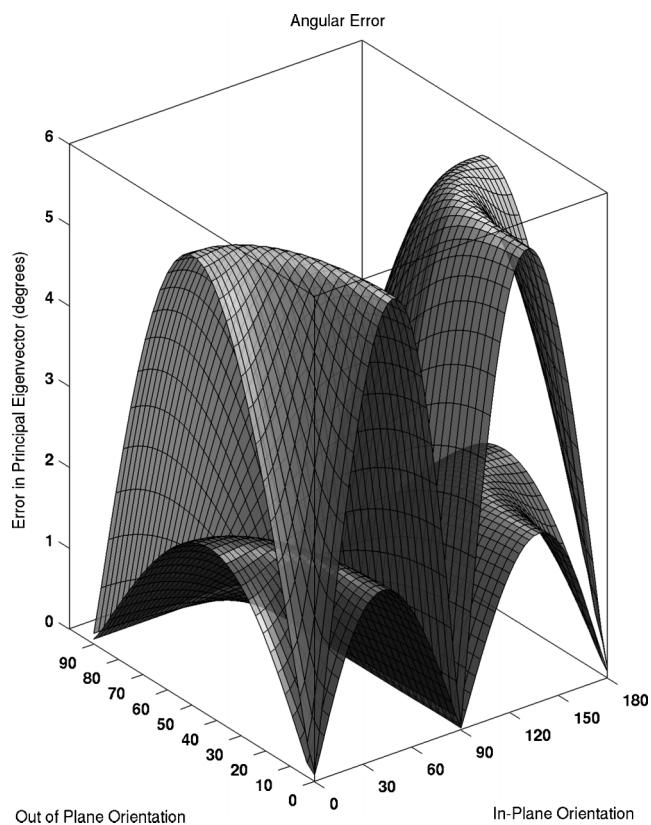


Figure 4. The consequences of neglecting to correct the signal intensity for changes in the volume of the voxel when correcting for eddy current induced distortions. The bias in the eigenvector orientation is plotted as a function of orientation (the horizontal axes show the in-plane and out-of-plane orientations in spherical co-ordinates), and anisotropy (the upper surface corresponds to a tensor with FA = 0.3, the lower surface corresponds to FA = 0.7).

typically, this is the number of unique orientations in which the diffusion encoding gradients are applied, since there is a benefit to be gained from using at least 20–30 unique orientations (24) for robust DT MRI (see below). Consequently, acquisition times are increasing and it is therefore increasingly likely (inevitable) that the participant will move during the scan, necessitating that all images are re-aligned before the tensor is estimated.

PITFALL 5: FAILURE TO ROTATE THE B MATRIX DURING MOTION CORRECTION

Correcting for subject motion is trivial. One does not expect magnifications or shears of the brain itself, and the translations and rotations are global. On first glance, one might naïvely assume that any off-the-shelf image registration package could be used to align the diffusion-weighted images to the first, non-diffusion weighted images, such as software used to realign functional MRI time series. However, one must keep in mind that unlike almost all other forms of medical imaging data, the DW images contain orientational information, and simply applying a rotation to the DW-images is not enough. To understand the problem, consider an extreme case of a diffusion encoding gradient being applied along the left-right axis of the head. In the body of the corpus callosum at the midline, diffusing water molecules are least hindered in this direction. Thus, the signal attenuation will be greatest and the diffusion-weighted image intensity will reach its minimal value. Now, in our thought

experiment, the participant rotates their head through 90 degrees, so that the midline callosal fibres are now perpendicular to the encoding gradient. In this case, we observe maximal hindrance to diffusion—and therefore the brightest DW signal intensity. If we blindly apply a rotation to this second data set to bring it into line with the first, without doing anything else, i.e. assuming that the encoding gradient is still applied in a left-right orientation, then we have a discrepancy. The signal intensities for the same voxel (midline corpus callosum) for the same gradient sampling vector are very different. The intensity in the second data set is ‘wrong’ and should be low rather than high. Clearly, if nothing is done about this, severe errors will result. Figure 5 shows how the error in the principal eigenvector varies for a single axial slice. These data were acquired from a normal healthy and motivated volunteer as described elsewhere (25) and still there are some voxels where the bias is as big as 4°, with the modal bias being around 1.5°. For less motivated volunteers who will move more during the acquisition, the bias will be much larger. It is also important to note that the bias is heterogeneously distributed and varies according to fibre orientation. If the fibre orientation is co-incident with the axis of rotation, there will be minimal effect [for example, a ‘nod’ has its axis of rotation about the left-right direction—and thus callosal fibres at the midline will be minimally affected, while those orientated perpendicular to the axis of rotation will be maximally biased (see Fig. 5)]. Thus, if uncorrected, one will again get biases that are heterogeneously distributed throughout the brain and dependent on fibre orientation. Although the step to remove this bias is trivial [i.e. applying the rotation to the encoding vectors (25)], it is not a universally adopted practice and is not provided in many popular

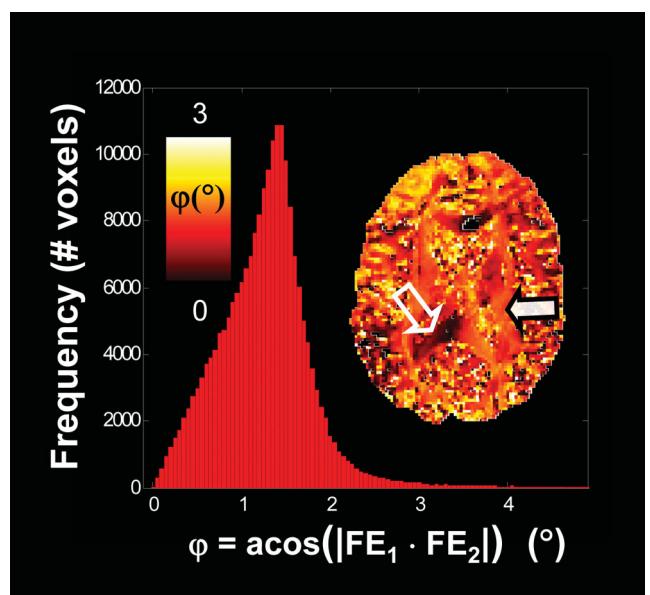


Figure 5. Illustration of the bias in eigenvector orientation if the orientations of the diffusion encoding vectors are not modified during correction for subject motion. The angular difference between the correct and estimated eigenvector, φ , is shown as a colour map. The filled arrow points to an area of large bias, and this corresponds to superior-inferior oriented fibres (oriented approximately perpendicularly to the axis of rotation). The empty arrow points to a region of low bias, corresponding to a region where the dominant fibre orientation is approximately parallel with the axis of rotation. The histogram shows the distribution of bias in eigenvector over the whole brain. (Adapted from Ref. (22))

software packages used for pre-processing of diffusion-weighted image data.

ESTIMATION OF THE DIFFUSION TENSOR

Once we have DW data that have been corrected for susceptibility and eddy-current induced distortions and subject motion, the next stage is to estimate the tensor in each voxel. Estimation of the diffusion tensor within each voxel requires the acquisition of images with diffusion-encoding gradients applied along non-collinear and non-coplanar directions (3,4). The diffusion-weighted signal intensity, DWI_m , obtained with the diffusion-encoding gradients applied along the m^{th} unit vector direction, $\mathbf{g}_m = [g_{x_m}, g_{y_m}, g_{z_m}]$, is given by

$$DWI_m = I_0 \exp(-\mathbf{B}_m \mathbf{D}) \quad [1]$$

where I_0 is the un-weighted signal intensity, \mathbf{D} is a vector containing the six unknown elements of the diffusion tensor:

$$\mathbf{D} = [D_{xx} \ D_{yy} \ D_{zz} \ D_{xy} \ D_{xz} \ D_{yz}]^T, \quad [2]$$

$$\mathbf{B}_m = b_m [g_{x_m}^2 \ g_{y_m}^2 \ g_{z_m}^2 \ 2g_{x_m}g_{y_m} \ 2g_{x_m}g_{z_m} \ 2g_{y_m}g_{z_m}], \quad [3]$$

and the diffusion weighting is given by B_m , in which the scalar variable b_m is solely a function of the magnitude and duration of the diffusion encoding gradients.

There are three widely used methods used for estimating the tensor in the literature. In order of increasing popularity and widespread usage, the first is nonlinear least squares (NLLS), second is weighted linear least squares (WLLS) and the most popular is ordinary least squares (OLS).

Non linear least squares estimation

In nonlinear least squares, the tensor is estimated directly from the signal of the form given in Eq. [1]. Due to the non-linear nature, iterative regression algorithms are needed to minimize the error between the predicted signal and observed signal intensities, the algorithm will search for changes in each of the unknown factors (the tensor elements) until the difference between observed and predicted signal intensity is within an acceptable tolerance. An example routine used for this is the Levenberg Marquardt algorithm (26). Given that the routine is iterative, the processing time is considerably longer than seen with the other two methods (OLS, WLLS). As discussed below, there are distinct advantages to using NLLS, despite the time penalty, but one has to be cautious to avoid another pitfall.

PITFALL 6: LOCAL MINIMA IN NON-LINEAR ESTIMATION OF THE TENSOR

A potential pitfall is landing in local minima and therefore getting biased estimates of the diffusion tensor. The potential of being subject to this pitfall is minimized by starting the search in the vicinity of the global minimum by obtaining an initial guess either through OLS or WLLS.

Weighted linear least squares (WLLS) estimation

In WLLS, a simple trick is done to take the estimation into a linear framework (3). If we take the logarithm of each side of Eq. [1], then we obtain:

$$\ln(DWI)_m = \ln(I_0) - \mathbf{B}_m \mathbf{D}, \quad [4]$$

If \mathbf{L} is an $N \times 1$ column vector containing the log of the diffusion-weighted intensities for each of N measurements, and \mathbf{B}

is a $N \times 7$ matrix, whose m^{th} row is given by:

$$\mathbf{B}_m = b_m [-g_{x_m}^2 \ -g_{y_m}^2 \ -g_{z_m}^2 \ -2g_{x_m}g_{y_m} \ -2g_{x_m}g_{z_m} \ -2g_{y_m}g_{z_m} + 1], \quad [5]$$

and \mathbf{P} is a vector containing all the unknown parameters

$$\mathbf{P} = [D_{xx} \ D_{yy} \ D_{zz} \ D_{xy} \ D_{xz} \ D_{yz} \ \ln(I_0)]^T, \quad [6]$$

then the entire set of measurements is conveniently represented as:

$$\mathbf{L} = \mathbf{B}\mathbf{P} \quad [7]$$

If we assume homoscedasticity, such that the variance in each DW signal is uniform—and independent of the sampling orientation, then \mathbf{D} can be found from the standard linear least squares solution

$$\mathbf{P} = (\mathbf{B}^T \Sigma^{-1} \mathbf{B})^{-1} \mathbf{B}^T \Sigma^{-1} \mathbf{L} \quad [8]$$

where Σ represents the covariance matrix of the data points that are fitted.

Ordinary linear least squares (OLS) estimation

If we make the assumption that the errors in our input data are independent and identically distributed (*i.i.d.*), then the covariance matrix in Eq. [8] becomes equal to the identity matrix multiplied by a scalar. Under these circumstances, the Weighted Linear Least Squares solution reduces to the Ordinary Least Squares Solution:

$$\mathbf{P} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{L} \quad [9]$$

The attraction of this solution is that it can be implemented extremely quickly. In fact, one can vectorise the solution for multiple voxels, speeding up computation efficiency even further. The ease of implementation and speed of solution explains why OLS is the most popular method for estimating the tensor and is the method implemented in popular tensor processing packages.

Note that if there are only seven measurements made—one at $b=0$, and six along non-collinear and non-coplanar direction, [such as with the dual-gradient scheme—(27)], then the matrix \mathbf{B} is square, and Eq. [9] trivially becomes: $\mathbf{P} = \mathbf{B}^{-1} \mathbf{L}$

PITFALL 7: HETEROSEDASTICITY IN THE FITTED DATA WHEN USING LINEAR REGRESSION

The first pitfall of analysis is the assumption that the noise in the fitted data is *i.i.d.* Under normal circumstances, we expect the variance in the diffusion-weighted signals at a particular voxel location to be homoscedastic. It is important, however, to realize, that what is being fitted is not the diffusion-weighted signals themselves, but the logarithm of the diffusion-weighted signals. Standard error propagation techniques (28) reveal that if the variance in a DWI signal is $\sigma_{DWI_m}^2$, then the variance in the log-transformed signal is:

$$\sigma_{\ln(DWI_m)}^2 = \frac{\sigma_{DWI_m}^2}{DWI_m^2} \quad [10]$$

One can see the immediate pitfall of assuming homoscedasticity. This assumption will only be valid for the case of an isotropic tensor where all the values of DWI_m are the same. The more anisotropic the tensor, the more the DWI_m s will vary with sampling direction, and the more incorrect the assumption will be. Thus, the error will be heterogeneously distributed throughout the brain.

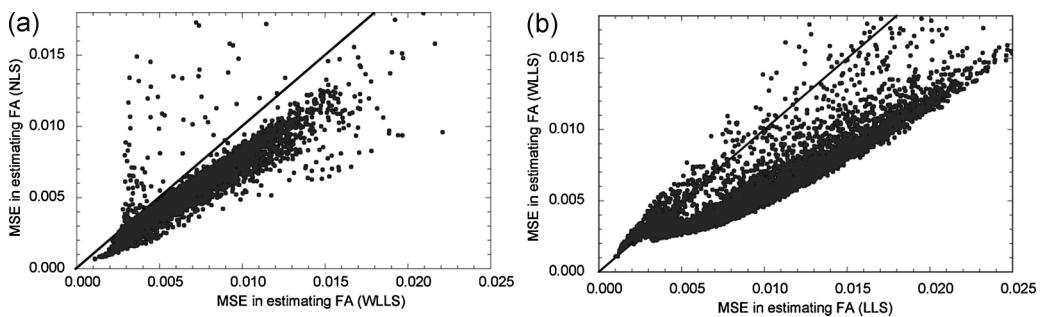


Figure 6. Effect of tensor estimation routine on the quality of resultant scalar data. The figure compares the mean squared error (MSE) in estimates of fractional anisotropy (FA) derived from 4000 Monte Carlo simulations of noisy diffusion data where the ground truth (the noise-free) tensor is known, for ordinary least squares (OLS), weighted linear least squares (WLS), and non-linear least squares (NLLS). The solid line at 45° in each plot is the line of identity. The authors are indebted to Dr Cheng Guan Koay, Section of Tissue Biophysics and Biomimetics, NICHD, National Institutes of Health for providing this figure.

PITFALL 8: DIFFERENT RESULTS BETWEEN DIFFERENT REGRESSION ALGORITHMS

Another pitfall comes when comparing tensor data when different estimation methods have been used. To demonstrate this, we are grateful to Dr Cheng Guan Koay for providing Figure 6. This shows the results of a simulation in which a whole human brain DTI data set was acquired and the tensor fitted using NLLS. The fitted tensor was then inserted into Eq. [1] above to generate a set of perfectly noise-free diffusion-weighted signal intensities. Random noise was subsequently added to these signals in quadrature, and the tensor estimated with each of the three estimation routines OLS, WLS, NLLS and the fractional anisotropy determined. This was repeated 4000 times (i.e. each time adding a new realization of the noise) and the mean squared error (MSE) in the FA determined for each voxel.

The figure shows that on the whole, the largest errors are found for OLS, while the smallest are for NLLS. Importantly, this also tells us that results computed with different regression algorithms cannot be directly compared. One therefore has to be cautious in multi-centre studies, for example, where each site may use a different software package to analyse their data. (One of the authors has direct experience of the same data set being sent out to two different centres for analysis and receiving back two different FA maps). Thus, one should ensure that the same package is used to analyse an entire data set in order for meaningful comparisons to be made.

FACTORS AFFECTING VOXEL WISE QUANTITATIVE PARAMETERS

Once the tensor is estimated, the usual practice is to derive the three eigenvalues (and associated eigenvectors) and use these to derive scalar indices, including the trace or mean diffusivity, and the anisotropy [most usually the fractional anisotropy, as it gives better SNR characteristics than other metrics such as RA and volume ratio, (29,30)]. While such parameters are quantitative and rotationally invariant, [and indeed tables of normative values appear in the literature—(31–33)] they are nevertheless, subject to biases which can confound analyses. This section covers issue affecting the mean and variance of particular parameters.

PITFALL 9: DEPENDENCE OF THE ESTIMATED MEAN DIFFUSIVITY ON THE B FACTOR

The first pitfall to consider in this section is that the trace of the tensor, although being a quantitative parameter is, in fact,

dependent on the amount of diffusion-weighting used to characterise it and there is no general agreement upon which b-value to use.

This dependence of the trace on the b-value is completely non-sensical if we take the true definition of trace being a scalar multiple of (three times) the mean square displacement per unit time of water molecules. If this is what we are measuring, then the b-value used should be irrelevant to the mean square displacement. The b-value dependence is attributable to the fact that diffusion tensor imaging models the three-dimensional displacement of water molecules as a multi-modal Gaussian function. However, in practice the diffusion of water molecules seen in the brain does not generally follow a Gaussian distribution, and in order to fully characterize the signal loss, one needs to invoke higher order cumulants (Taylor series coefficients) in the signal attenuation expression. In diffusion tensor imaging, these higher order terms are considered as negligible and the lower the b-value, the smaller these higher order terms become in relation to the diffusion tensor's contribution to signal attenuation—and thus, the log of the signal does appear to scale linearly with the b-value.

At b-values of around 1000 s mm^{-2} , the trace is remarkably uniform throughout parenchyma (34), irrespective of the huge variations in anisotropy. While in grey matter, the three eigenvalues will be approximately equal, in white matter, the eigenvalues will be very different. For example, when $\text{FA} = 0.7071$, if the tensor is prolate and cylindrically symmetrical (such that the two smallest eigenvalues are equal), the largest eigenvalue is twice as big as the mean diffusivity, while the smallest is only half as big. Nevertheless, at b-values of around 1000 s mm^{-2} , the sum of the three eigenvalues always gives approximately the same answer. This phenomenological observation means that the variance of values from within an ROI will be minimally affected by having different tissue types contained within the ROI (e.g. grey matter and highly anisotropic white matter). One can therefore potentially use any ROI on an intra-subject basis to act as an internal control to determine whether a particular area has abnormal diffusivity. However, the issue is that there is no universal agreement on what b-value should be used for diffusion tensor MRI experiments. Figure 7 shows what happens when the b-value is changed. As the diffusion-weighting is increased, there is an increasing dissociation between white and grey matter (35), with the trace in the white matter being lower than in grey matter. This not only

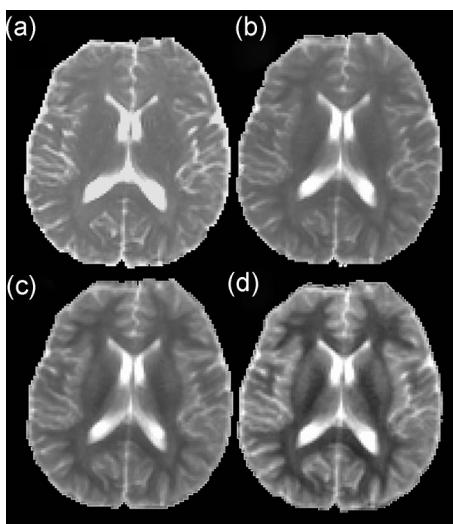


Figure 7. Variation of the mean diffusivity as a function of b-value. (a) 1000 s mm^{-2} ; (b) 2000 s mm^{-2} ; (c) 3000 s mm^{-2} ; (d) 4000 s mm^{-2} . All four images are windowed to the same extent. Note that as the b-value is increased, the dissociation between mean diffusivity in the white and grey matter gets progressively larger.

introduces more heterogeneity within ROIs, but also means that comparison with normative databases or data from other centres and, therefore, multi-centre studies, are going to be problematic unless the same degree of diffusion-weighting is employed. The take home message is that within an ROI, the mean and variance of the trace will be dependent on the b-value.

PITFALL 10: NOISE INDUCED POSITIVE BIAS IN ANISOTROPY DUE TO EIGENVALUE REPULSION

The next pitfall concerns anisotropy measurements. The effect of noise on anisotropy measurements derived from DT MRI is now well known—and was first described by Pierpaoli and Basser (36). In short, noise in the diffusion-weighted signals will mean that, even in a perfectly isotropic medium—such as a pot of water, one will never get three identical eigenvalues. The noisier the signals, the bigger the discrepancy of the eigenvalues will be. Thus, there is a noise-induced bias in measurements of anisotropy. However, this becomes less marked as the anisotropy increases. Figure 8 illustrates the consequences of this. In this figure, a DTI data set was collected and the tensor fitted in each voxel. The noise-free diffusion-weighted intensities were derived, and then complex noise added to generate DT MRI data sets with progressively noisier data. The plot shows how the mean and variance of the anisotropy value changes. In both cases, the variance in the anisotropy increases as the added noise increases. However, the mean value remains approximately constant in the white matter, but increases rapidly in the grey matter (36) (See Fig. 8). Thus, once again—unless the acquisitions are matched, so that the signal to noise ratio in the non-diffusion weighted images is the same and, equally importantly, so that the same number of $b = 0$ images and diffusion-weighted images are acquired, then comparing anisotropy values across different subjects/time-points/centres is particularly problematic.

PITFALL 11: NOISE INDUCED NEGATIVE BIAS IN ANISOTROPY AND MEAN DIFFUSIVITY DUE TO RICIAN NOISE DISTRIBUTION

Another potential issue concerns datasets with low SNR in the diffusion-weighted scans. This can arise from data with

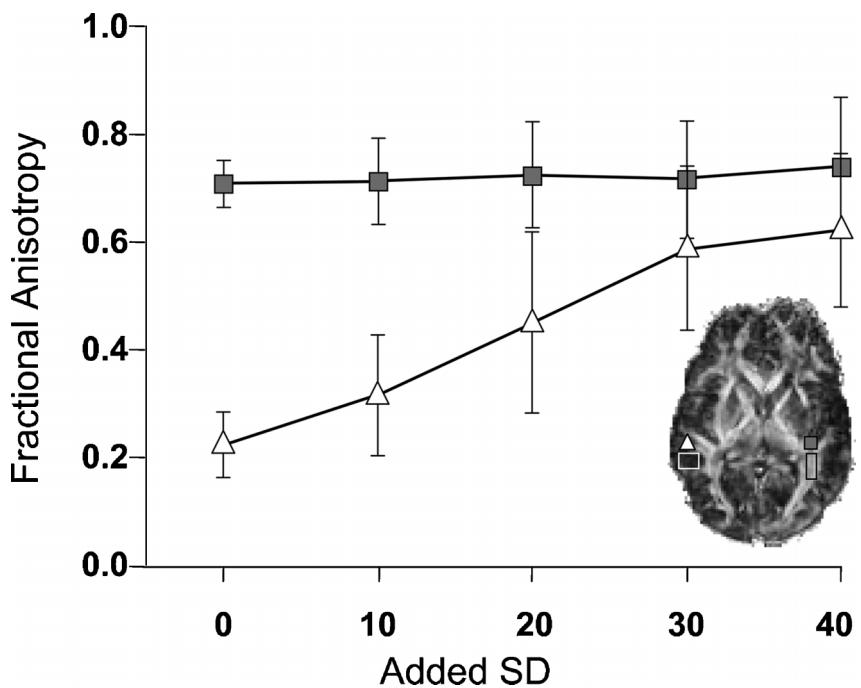


Figure 8. Simulation of the effects of noise on estimated fractional anisotropy. This simulation began by fitting a tensor to each voxel and then, for a given set of encoding vector directions and amplitude, deriving the noise-free diffusion-weighted intensities. Noise was then added in quadrature and the tensor re-estimated in each voxel, and the FA determined. Results are shown for two ROIs, the first (filled grey squares) was placed in anisotropic white matter, while the second (white filled triangles) was placed in a region of grey matter (low anisotropy). The error bars show 1 standard deviation (SD). Note that for both white and grey matter, the error bars increase with the added noise. In white matter, the mean value remains almost immune to variation in added noise, whereas the mean value in grey matter increases rapidly. The authors are indebted to Prof Geoff Parker, Manchester University, UK, for supplying a form of this figure and allowing us to include it here.

insufficient SNR in the $b=0$ scans, and moderate b -values, or could also arise in data with high SNR in the $b=0$ scans, but in which large b -values are used. The issue concerns the Rician distribution of the noise (37)—and the presence of the rectified noise floor (38,39). This leads to an effective damping of signal attenuation which, in turn, leads to an underestimation of diffusivity (38). Again, in anisotropic media, this effect will be orientationally dependent. In a tensor with high anisotropy, the diffusivity is much higher along some axes than others, so the underestimation of diffusivity will only occur along a limited number of orientations. This will, in turn, lead to a negative bias in anisotropy, and in the trace of the tensor (39). Moreover, as the underestimation of the trace becomes more marked at higher anisotropy, one can find artificial correlations between trace and anisotropy (39)—which are normally assumed to be independent metrics in statistical analyses. If this is an issue, one can use alternative analyses that account for rectified noise floor (38,39). It has also been found that even within the three popular regression methods (OLS, WLLS, NLLS) that do not explicitly account for the noise-floor, there are differential responses with NLLS outperforming WLLS which outperforms OLS (39). So, again, it is important when comparing results to establish not only what acquisition was used—but also what tensor estimation routine was used.

PITFALL 12: ORIENTATIONAL DEPENDENCE OF VARIANCE IN DT ESTIMATES DUE TO CHOICE OF ENCODING VECTORS

Without due care and attention to the design of the acquisition scheme used to acquire the diffusion-weighted data, there can be another pitfall when comparing indices derived from DT-MRI. If the number of sampling orientations is low and not uniformly distributed over the surface of a sphere, then one finds that the variance in derived indices can be strongly dependent on the orientation of the structure (24). One finds, for example, that the lowest variance in a parameter such as FA is found when the fibre is aligned with one of the sampling orientations, and is largest when the fibre is at the greatest angle to the sampling vectors. The orientational dependence becomes more marked as the anisotropy increases. This means that the statistical power to detect a change/group difference will depend both on the anisotropy and on the orientation of the structure with respect to the magnet. Moreover, within the same brain, different structures are oriented differently and have different anisotropies—so the power to perform intra-subject comparisons is spatially heterogeneous. The uncertainty in fibre orientation (40) is particularly sensitive to this effect, and thus extra care needs to be taken when performing statistical comparisons of orientational information derived from DT MRI (41,42). This effect can be minimized by increasing the number of unique sampling orientations (typically at least 30 unique and uniformly distributed directions will suffice (24)—but time pressures often mean that this is impractical, so this potential pitfall in statistical analysis must be borne in mind.

EXTRACTING QUANTITATIVE PARAMETERS

Once the parametric maps of interest have been computed, it may be necessary to extract summary measures, in a format suitable for a statistical comparison, from either the whole brain or specific anatomical locations. For this purpose three alternative approaches are available: region of interest (ROI), histogram and

voxel-based (VB) analysis. This section reviews the potential sources of error hidden behind ROI and histogram analysis, while pitfalls associated with VB analysis are discussed in the Inter-Subject Comparison section.

Region of interest analysis

A simple approach to extracting DT MRI parameters representative for a specific area of the brain is to specify a region of interest (ROI) and extract the mean parameter of interest from the voxels within it. An ROI is typically outlined and positioned manually on a suitable co-registered image.

PITFALL 13: BIAS INDUCED BY PARAMETER VALUE ON ROI POSITIONING PERFORMED DIRECTLY ON PARAMETRIC MAPS

If ROIs are defined directly on the parametric map of interest (e.g. FA), the intensity on the map might spuriously influence the position of the ROI boundaries. For example, in the case of FA, a rater can be biased towards drawing the boundaries of an ROI around those voxels characterised by higher anisotropy. Ultimately, this could lead to artificially removal of group differences. To avoid this type of bias, it is customary to use an image where the contrast is independent of the quantity to be measured, typically a 'conventional' anatomical image (T_1 or T_2 weighted). Due to geometric distortions, however, achieving a good match between DT MRI and anatomical images can be problematic, as previously discussed. If EPI distortions cannot be compensated for, or alternative anatomical images are not available, a reference for ROI positioning must be chosen among DT MRI data, and the most obvious choice for an 'independent' image is an image acquired with a diffusion-weighting close to zero ($b=0$ image). These images have the same susceptibility induced distortions as the DW images, but a contrast that is independent of diffusion.

PITFALL 14: EFFECT ON ROI POSITIONING OF SUBTLE MIS-REGISTRATION BETWEEN B_0 AND DW DATA

Even after image realignment, the highly localised differences in image contrast between DW images and $b=0$ images can result in subtle registration errors, and ultimately in fine misalignment between $b=0$ images and DT maps. This effect can be appreciated when positioning small ROIs in subcortical structures, as shown in Figure 9. Under these circumstances, even a small error in ROI placement can have a large effect on the mean FA derived from the ROI, due to the rapidly changing values in FA maps between neighbouring tissues. The best strategy for positioning ROIs might vary depending on the shape, the size and the location of the region. Before extracting the values of interest, the rater is encouraged to check the position of the ROI on the image of interest.

Histogram analysis

Another popular method of analysis of quantitative MRI image is histogram analysis. A histogram is a frequency distribution showing the proportion of voxels in an image within a given range of values of the parameter of interest (e.g. FA, MD, etc.). Usually some numerical features (histogram metrics) are selected to characterise the shape of the graph; the metrics are extracted for each subject in the study, and a statistical comparison between patients' and controls' average metrics are performed to look for any group differences. Histogram analyses reflect global changes, testing across the whole brain and they

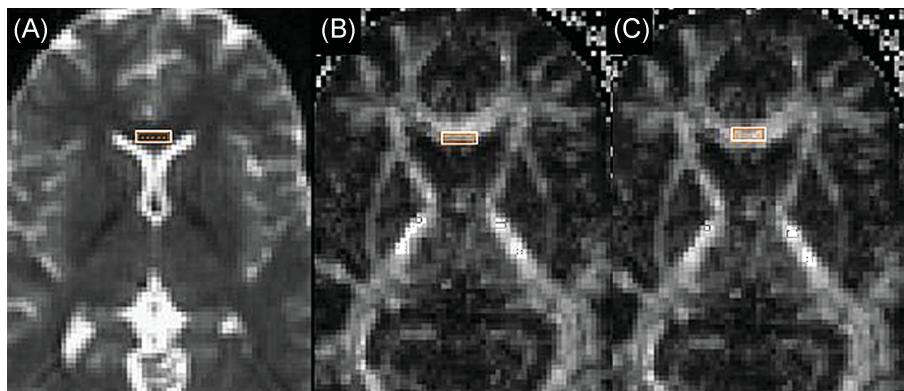


Figure 9. The effect of local misalignment between DW images. A rater was asked to position a rectangular ROI in the genu of the corpus callosum, first on a $b = 0$ image (A), then on a FA image (C). The central panel shows the same ROI as in A, but overlaid onto the FA image. The mean ROI FA is 0.63 for the ROI originally positioned on the $b = 0$, and 0.73 for the ROI positioned directly on FA.

are thus particularly indicated when dealing with a diffuse disease.

PITFALL 15: CSF CONTAMINATION IN HISTOGRAM ANALYSES OF DT MRI DATA

Histogram creation requires the removal of the tissue of no interest (typically CSF). Ideally, image segmentation should be carried out on an image whose contrast is independent of the quantity to be measured, but co-registered with the parametric map under study. However, as with ROI placement, geometric distortions can impair the registration between EPI data and high-resolution, anatomical images. Moreover, the segmentation process is complicated by the inevitable presence of some RF inhomogeneity. Due to the low resolution of DT MRI data, partial volume effects are expected to be substantial, and therefore whole brain histograms will always have a contribution from CSF. This confound may be particularly evident when the subjects in one group are more likely to be atrophic than those in the other group. In this case poor segmentation of CSF might result in different degrees of partial volume for the two groups under comparison, with brain atrophy having an undefined contribution to any observed change. It is possible to adjust the statistical analysis for these effects by including an independent measure of brain atrophy as a confounding factor in the analysis. For example, in a study of patients with relapsing-remitting multiple sclerosis (43), the brain parenchymal volume (whole brain volume after CSF removal) was found to have a significant effect on all FA histogram metrics, despite there being no significant difference in the mean value of this quantity between patients and controls. It should be noted, however, that in most histogram analyses of diffusion MR data such corrections are rarely performed.

Dealing with partial volume

Partial volume between differing tissue types, particularly with CSF, can be a problem also when dealing with analysis methods other than histogram. The mean diffusivity of water in CSF is up to 3 times larger than in brain parenchyma and therefore any signal contamination by CSF can result in a severe bias of the estimated mean diffusivity for other classes of tissue. This is particularly relevant for the cortex whose thickness is approximately 2–3 mm, bearing in mind that the typical voxel resolution in DT-MRI is also on the order of 2–3 mm.

PITFALL 16: CSF CONTAMINATION IN TRACT-SPECIFIC MEASUREMENTS

There is an increasing trend to conduct ROI-type studies by using a fibre-tracking algorithm to define a particular tract, and then to extract the mean value of the parameter of interest (e.g. trace, anisotropy, eigenvalues) from within the tract (e.g. (43–46)). While this approach is attractive in that it provides for a more objective approach to ROI delineation, and allows one to extract data from ROIs that would otherwise be extremely cumbersome to draw by hand, this approach is also not free of the pitfalls of partial volume contamination. One can readily check for such effects by colour-encoding the reconstructed tract by the mean diffusivity (47). As described earlier, at b -values of around 1000s mm^{-2} , the mean diffusivity is fairly uniform across parenchyma, thus any partial volume with the CSF will be immediately apparent. Figure 10 shows tracking results obtained in the body of the corpus callosum and fornix of a normal healthy volunteer. Note that in the majority of the callosal fibres, the trace is uniform and remains low. In the fornix, however, there are lots of ‘hot spots’ where the trace is substantially higher. Thus, although the *anatomy* of the tract reconstruction looks reliable, one has to bear in mind that pathways that pass close to CSF interfaces will be equally impacted by partial volume contamination, leading once again to spatial heterogeneity in biases in anisotropy and mean diffusivity (both in terms of the mean values and the standard deviations).

CSF suppression during acquisition can mitigate the problem of partial volume (48,49). A reproducibility study showed that with fluid-attenuated inversion recovery (FLAIR) DW MRI the apparent diffusion coefficient (ADC) histogram is shifted to lower values with a significantly narrower peak (50). Moreover, it has been shown that fibre tracking results improve when FLAIR acquisitions are employed (51–53). However, it should be noted that FLAIR acquisitions result in lower signal-to-noise ratio than non FLAIR acquisitions, they require longer scan time, and they are incompatible with cardiac gating. An alternative solution, originally proposed by Pierpaoli and Jones (54) is the use of two-compartment tensor model, where one compartment to be fitted has the properties of CSF (isotropic, mean diffusivity = $3.0 \times 10^{-3} \text{ mm}^2\text{s}^{-1}$), and the other compartment is modelled freely as a tensor. The contribution to the diffusion-weighted signal from the CSF component falls into the first component,

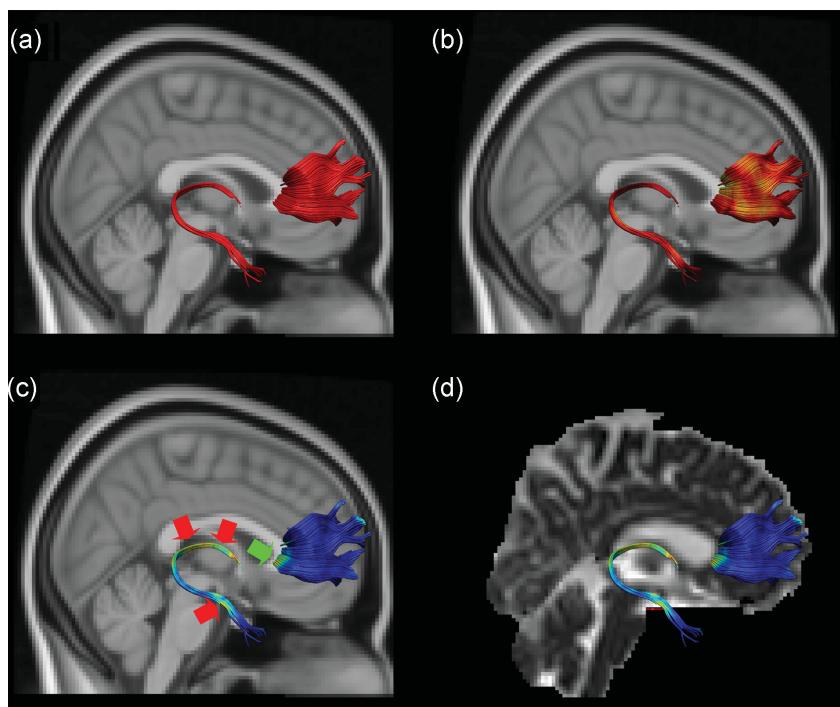


Figure 10. Illustration of the problem of partial volume contamination in tract-specific measurements of anisotropy. The figure shows tracking results from two seed regions—one in the fornix and one in the genu of the corpus callosum. In a–c, the ‘backdrop’ is an atlas SPGR image, whereas in d, the backdrop is the non-diffusion weighted ($b = 0 \text{ s mm}^2$) image. In (a), the typical deterministic tracking representation is shown—with the tracts represented in a single colour. The reconstructions are smooth and consistent with known anatomy—so there is nothing to suggest that there may be a problem. In b, the tracts are coloured according to the fractional anisotropy, revealing that the anisotropy of the callosal fibres is higher than those in the fornix. (The mean FA along the genu fibres is 0.48 ± 0.19 , and along the fornix fibres it is 0.33 ± 0.11 , suggesting that the callosal fibres may be more ‘organized’. However, when the tracts are encoded for mean diffusivity (c), one sees that the trace is much higher in the fornix than in the genu. The arrows point to ‘hotspots’. As stated elsewhere, the mean diffusivity in parenchyma has been shown to be largely independent of anisotropy—which is the case in most of the genu fibres, where there is minimal variation in colour-coding. Close inspection of (d) shows that the areas of hotspots correspond to regions where the fibres run close to CSF-filled spaces, (red arrows for fornix, green arrow for genu) and thus we are visualizing partial volume contamination which in turn leads to a reduction on the estimated anisotropy and elevation of mean diffusivity (fornix = $1.2 \pm 0.3 \times 10^{-3} \text{ mm}^2 \text{s}^{-1}$; genu = $0.8 \pm 0.1 \times 10^{-3} \text{ mm}^2 \text{s}^{-1}$)

with the idea being that the second component is less affected by CSF contamination. This idea has recently been refined by Pasternak *et al.* (55) and seems a promising way forward to address partial volume issues in the *analysis* rather than in the acquisition pipeline.

INTER-SUBJECT COMPARISON

Most studies based on the use of DT MRI for clinical/neuroscience research set out to compare DT MRI parameters in two or more groups of subjects, or to assess the association between DT MRI parameters and clinical/performance variables. Voxel-based (VB) analyses, in particular, are becoming more and more popular as a method of analysing quantitative images because such methods are highly automated, and require minimal intervention from the user. Originally derived for structural (T1-weighted) data (56), they are becoming increasingly common for analysis of DT MRI data. They rely on the accurate alignment of images from different subjects onto a common template, to achieve a correspondence between a particular voxel position in each image and the same anatomical structure across subjects. Voxel-wise statistics are then carried out across the whole brain,

removing the need for making an *a priori* spatial selection and hypothesis. In this section, we highlight some pitfalls that may influence the analysis in a non-trivial fashion. Some of the pitfalls here reported are specific to VB analysis of FA, due to the ‘unique’ contrast of these images, while others (e.g. pitfalls 17, 19, 23) are relevant for all DT MRI data, including mean diffusivity and eigenvalues. It should also be mentioned that several well known problems of VB methods in general, are relevant also for the analysis of DT MRI data. We do not discuss these problems in detail, as these have been covered extensively by others (e.g. (57–59)).

Normalisation of DT MRI data

All VB analysis methods (regardless of the type of images to be tested) rely on the accurate alignment of images of different subjects, and the normalisation step is thus critical, with registration errors having an undefined contribution to the result of the comparison (57–59). When dealing with DT MRI, the registration step is even more challenging, and several aspects must be considered.

The first question to address is which images should be used to drive the image registration. An option is using the $b = 0$ images

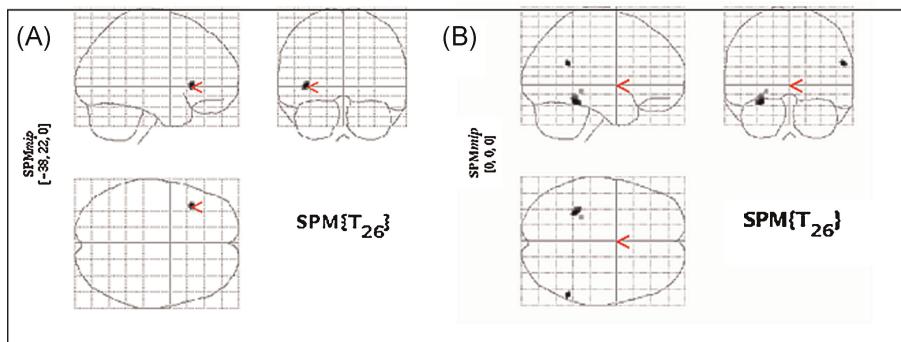


Figure 11. Results of the voxel based comparison of FA images between 14 healthy subjects and 14 patients with schizophrenia. The black spot indicates position and size within a glass brain of the areas of reduced FA in patients when using images normalized with affine transformations only (A) and with non-linear normalization (B). Both results are consistent with previous literature.

to drive the registration either onto an EPI template or onto a T₁-weighted, high resolution template. This is the simplest approach, but it is not ideal. Due to the poor contrast in b=0 images and since no orientational information is considered during the warping procedure, this method may introduce white matter alignment errors (60). A general problem with VB approaches is that a local misalignment between images can be erroneously interpreted as a signal change. When dealing with DT MRI, this problem is further complicated by the possibility of achieving a successful alignment of topological characteristics (overall brain shape, tissue boundaries, etc.) without an exact alignment of white matter tracts. A better result can be achieved when using FA maps to drive registration (60). An alternative solution to using a single scalar image to drive registration is the use of all 6 components of the diffusion tensor (61). Multi-channel registration algorithms are more expensive in terms of computational time than single-channel ones, but their performance was shown to be superior (60). In cases of registering non-scalar data like this, as already discussed with respect to motion correction, any rotation in the image position requires a further rotation of the orientational information represented at every voxel position (62), or unpredictable errors may occur.

PITFALL 17: AFFINE VS NON-LINEAR NORMALIZATION IN VOXEL-BASED ANALYSES

VB analysis is potentially highly reproducible. In practice, however, the number of options available for pre-processing the data and setting up the statistical comparison yields a huge variety in the results obtained by different groups in similar experiments. The first factor which may affect the outcome is the choice between affine and non-linear normalisation. Due to the local differences in shape and size of brain structures, linear registration only provides a gross alignment between brains. This first alignment is usually (but not always) followed by a non-linear registration step (also known as warping). One might think that skipping the non-linear step would simply reduce the sensitivity of the analysis. In order to test this hypothesis we used VB-analysis to analyse twice the same FA data, obtained from a group of 14 schizophrenic patients and a group of 14 age, gender and IQ-matched healthy controls. First we performed a voxel-wise comparison of the mean FA images after affine normalisation only. This analysis revealed a cluster of reduced FA in patients close to the left uncinate fasciculus (which is entirely consistent with previous literature reports, e.g. ref. (63)). Then the same analysis was repeated on the images after non-linear normalisation (but otherwise identical methodology). The difference in

the uncinate fasciculus disappeared, while two other clusters appeared in the left cingulum and the right superior longitudinal fasciculus, which have also been implicated in previous studies of schizophrenia (ref. (64) and ref. (45), respectively). These results are summarised in Figure 11. So, although all these findings are consistent with symptoms of the disease, this simple and single modification in the processing pipeline leads to remarkably different conclusions.

PITFALL 18: EXCESSIVE/ERRONEOUS WARPING OF FA IMAGES

The second pitfall concerns non-linear normalisation. The most common approach to non-linear registration is based on the use of a linear combination of basis functions (e.g. polynomials, sinusoidal functions, etc.). Although excellent registration algorithms have become available, warping can go badly wrong when dealing with images of unusual contrast and relatively gross resolution, such as those taken from DT MRI data. For this reason, it is a good idea to constrain the amount of warping by some kind of regularisation (i.e. mathematical methods of imposing stability on an ill-posed problem, often by incorporating some *a priori* information). Despite regularisation, however, results similar to those shown in Figure 12 are frequently observed.

Smoothing

Smoothing is achieved by spatial filtering with (usually) a 3D Gaussian kernel. It is considered an essential step prior to statistical comparison in order to achieve three main goals:

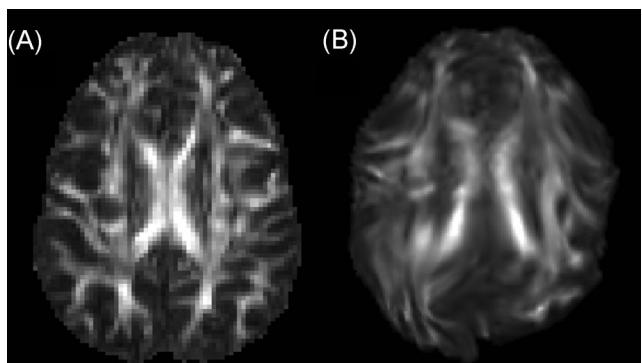


Figure 12. FA image before (A) and after (B) warping to match a standard FA template.

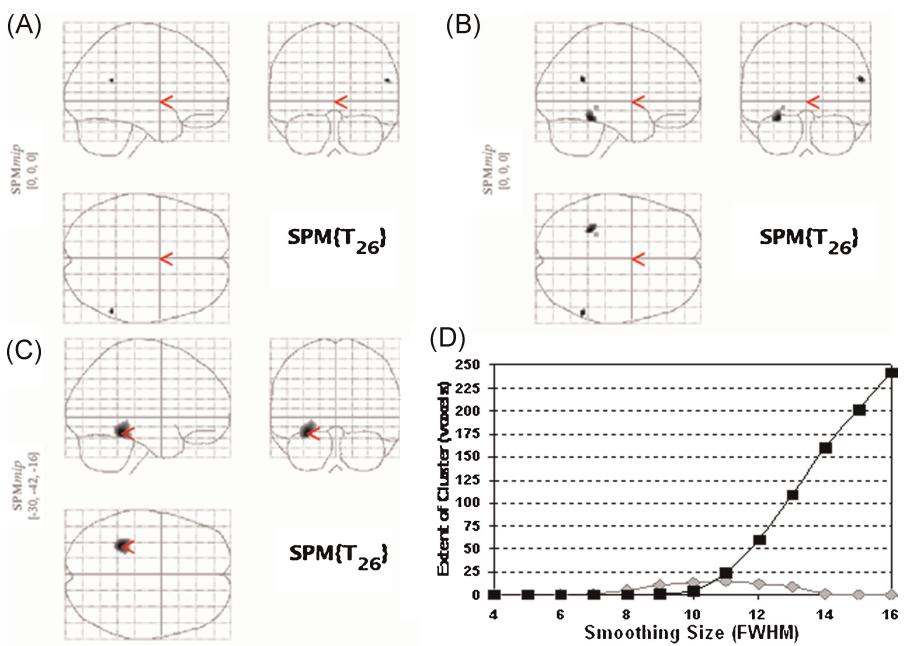


Figure 13. The effect of smoothing on the results of VB analysis. The location and size of the clusters of reduced FA found in patients with schizophrenia compared to healthy controls changes with smoothing kernel. The figure shows the results for FWHM of 8 mm (A), 12 mm (B) and 16 mm (C). The graph in D shows the size of the cluster measured in number of voxels as a function of the smoothing kernel for the two clusters: one centered at MNI coordinates [54, -50, 22] (grey diamonds) and the other centered at MNI coordinates [-28 -44 -18] (adapted from ref. 68).

increasing the signal-to-noise ratio by averaging, compensating for residual misalignment, and making the data more normally distributed which is a pre-requisite for parametric statistics [which assumes the residuals to be normally distributed, i.e. Gaussian (65,66)]. The width of the filter is usually defined by the full-width-at-half-maximum height (FWHM) of the Gaussian filter.

PITFALL 19: CHOICE OF SMOOTHING KERNEL AND THE MATCHED FILTER THEOREM

Smoothing tailors the search to the spatial extent of the filter, through the Matched Filter Theorem (67). This means that the width of the filter used to process the data sensitises the analyses to differences of spatial extent comparable with the filter width itself. Thus, when using a single low-pass filter to pre-process DT MRI data prior to performing a voxel-wise comparison, one should specify *a priori* the spatial extent of the expected effect. This theoretical observation was confirmed empirically by Jones *et al.* (68), who repeated the VB comparison of FA between a group of patients with schizophrenia and a group of healthy controls varying the size of the smoothing kernel. They showed that the detectability and the extent of the two areas of abnormality identified by VB analysis (located in the right superior temporal gyrus and at the boundary between temporal lobe and cerebellum, respectively) were indeed a function of the filter size (Fig. 13), thus stressing the importance of recognizing the effect of the Matched Filter Theorem on VB analyses of DT MRI data.

PITFALL 20: MISLOCALISATION OF GROUP DIFFERENCES IN HETEROSEDASTIC DATA DUE TO SMOOTHING

The second pitfall, related to the previous one, concerns the increase in the amount of partial volume caused by smoothing. A consequence of this partial volume effect is that it is quite common for VB analysis to report changes in FA at locations away from main white matter bundles. This questions the interpretability of these changes and poses serious doubts on the

applicability of these methods of analysis to DT MRI data. Clearly, the larger the smoothing kernel the more the partial volume artefact will be. No systematic method can be advocated for selecting the size of the smoothing kernel, although a 'rule of thumb', originally determined for a similar analysis of positron emission tomography (PET) and functional MRI data, states that the smoothing kernel should be at least two or three times the voxel size of the images to be compared. Given the typical voxel size for DT MRI is approximately 2.5–3 mm³, according to this rule a smoothing kernel of 8–10 mm³ should be appropriate. Looking at Figure 14, however, it becomes clear that any power to differentiate between anatomical structures is lost on FA images between 8 and 12 mm³ smoothing (the latter being a very popular choice for VB analysis of FA images), regardless of the type of normalisation.

The consequences of smoothing unfortunately go beyond simple 'blurring', as this procedure introduces a systematic bias in the anatomical localisation of group differences (57). If a group difference occurs in a brain area at the boundary between regions of low and high across-subject variance, then smoothing will artificially 'move' the difference towards the region of low variance. This is simply a consequence of the higher sensitivity for detecting differences in the mean signal intensity in areas where the variance is small, combined with the property of smoothing of 'smearing out' the difference around the real location. Here, we have simulated this effect and its dependency on the width of the smoothing filter (based on code provided by John Ashburner, Wellcome Trust Centre for Neuroimaging, University College London at:

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind04&L=SPM&P=R32206>). The matrix in Figure 15 simulates the signal intensity in 100 voxels for 100 different subjects. The mean signal is zero everywhere, but the variance in the first 50 voxels is 10 times the variance in the last 50 voxels. If we add an offset of 100 to the mean signal at the interface only for the first 50 subjects,

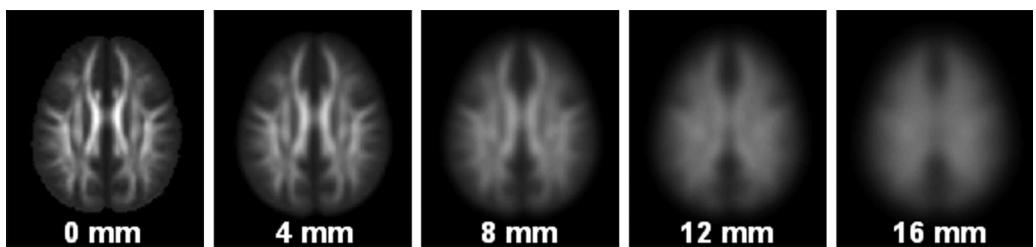


Figure 14. The effect of smoothing on anatomical resolution. The same image was filtered with increasingly large smoothing kernels, ranging from 0–16 mm. The picture shows that for smoothing kernels larger than 10 mm, any anatomical specificity, in terms of white matter structure, is lost (e.g. the cingulum, clearly visible for smoothing kernels ranging from 0–8, is merged with adjacent structures in the following images).

this will simulate a group difference (with the two groups made of 50 subjects each). We smooth the data with smoothing kernels ranging from 5 to 15 FWHM. If we run a *t*-test, and plot the *t* statistic as a function of the voxel position, we expect a maximum at voxel position 50.5. On the contrary, what we observe is a progressive shift of the maximum peak towards the low-variance areas as the smoothing kernel increases.

A knee-jerk reaction might be that situations of an abrupt change with spatial position of the across subject variance by a factor of 10 are unlikely to be encountered in real data, but with images like maps of FA, characterised by great variations in signal intensity between neighbouring tissues, a sharp increase in the variance is a frequent occurrence. The variance of FA is indeed very spatially heterogeneous, as demonstrated

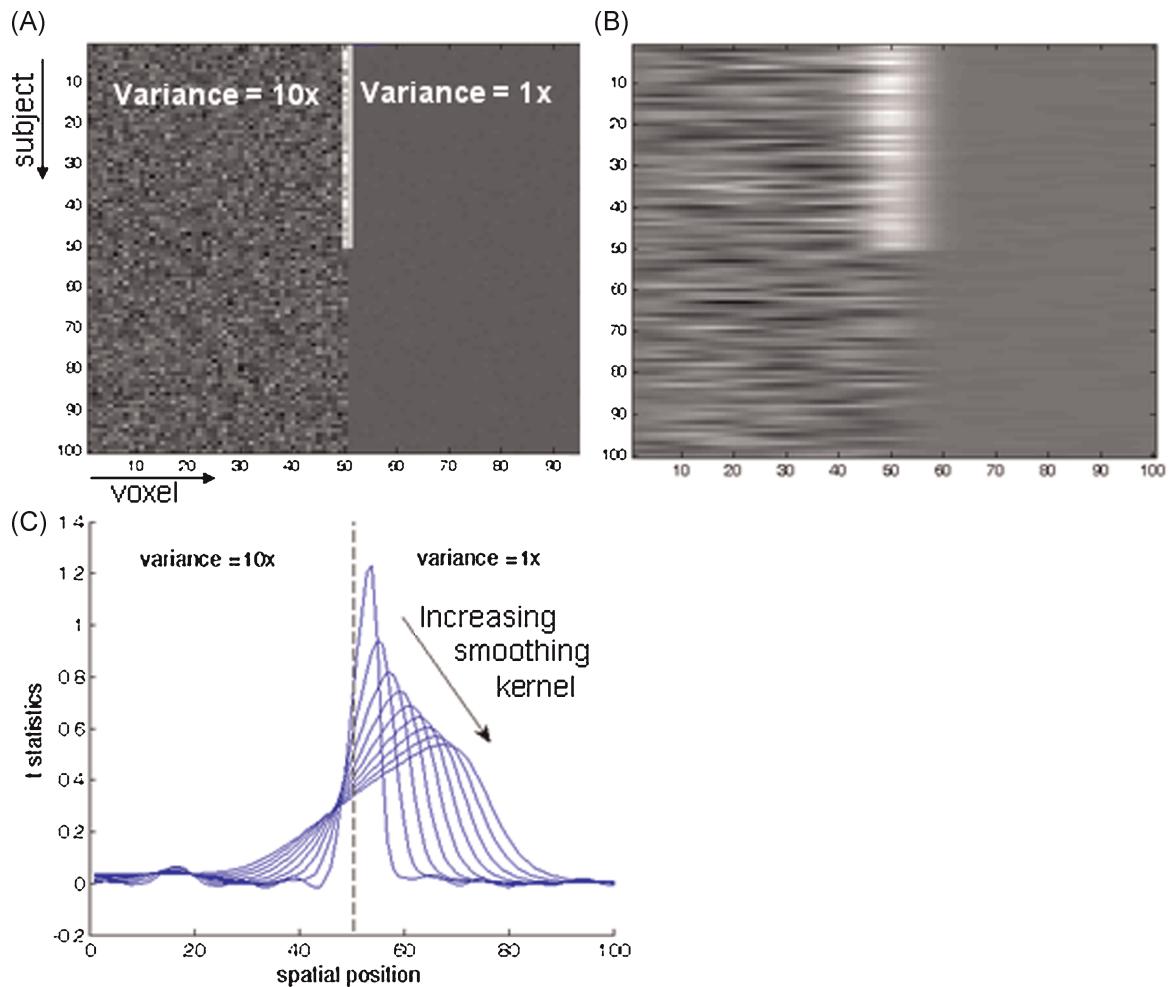


Figure 15. Results of the simulation used to illustrate pitfall 20. The simulation is based on code provided at <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind04&L=SPM&P=R32206>. The matrix in A simulates the signal intensity in 100 voxels for 100 different subjects, with signal zero mean, and variance in the first 50 voxels that is 10 times the variance in the last 50 voxels. An offset of 100 was added to the mean signal at the interface only for the first 50 subjects. The matrix in B shows the same data after smoothing with FWHM of 8 mm. Panel C shows the *t* statistics as a function of the voxel position for differing smoothing kernels. See text for further details.

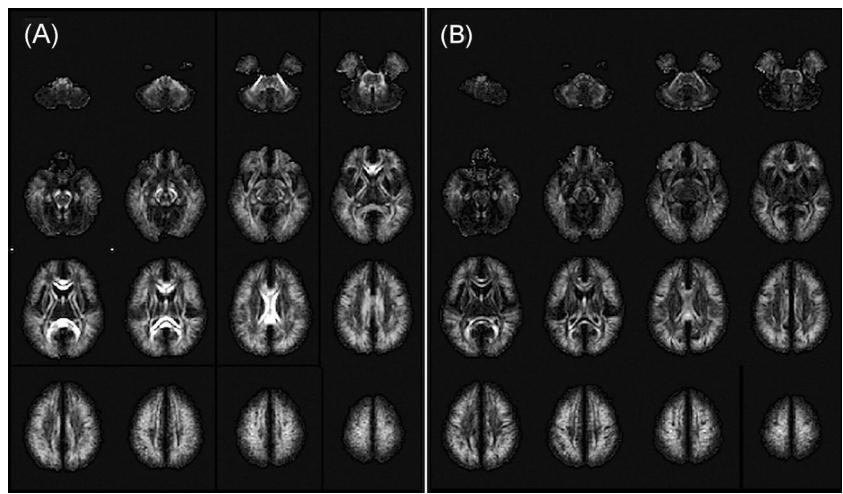


Figure 16. Images of the variance in FA images across 43 subjects, showing the great spatial heterogeneity. Images were affine normalised for panel A and non-linearly normalised for panel B.

by Figure 16, which displays FA variance across 43 healthy subjects.

PITFALL 21: SPATIALLY HETEROGENEOUS SENSITIVITY TO GROUP DIFFERENCES

This spatial heterogeneity in variance of FA maps is also expected to impact the sensitivity of VB analysis to group differences according to their anatomical location. It is implicitly assumed that the spatial power to detect a group effect is uniform throughout the brain—but as shown in Figure 16—this is clearly not the case. Consequently, one may have a *global* group effect—but, due to the spatial heterogeneity in the ability to detect that group difference, the VB analysis would reveal distinct clustered regions of group differences. Note that when moving from affine to non-linear registration, the spatial variability is reduced but is by no means fully removed.

Statistical analysis

In VB methods, the statistical analysis is performed on a voxel-by-voxel basis, usually based on multiple regression or equivalent methods (66), followed by some test of significance. Most image analysis packages use parametric tests, although non-parametric methods are also available.

PITFALL 22: NON-NORMALLY DISTRIBUTED RESIDUALS

Parametric statistics require the residuals to the model to be normally distributed. Here we address the question whether this hypothesis holds true for FA images. Normality can be tested using the Lilliefors Test (69), an adaptation of the Kolmogorov-Smirnov goodness-of-fit test. It tests the null hypothesis that data come from a normally distributed population, by comparing the empirical distribution of residuals with the cumulative distribution function of the normal distribution. The null hypothesis is rejected when the difference between these two distributions is large enough to be considered significant. The residuals can be computed from the same population of 43 subjects used to estimate the variance in Figure 16. If we set an alpha value of 0.05, we can test the distribution of FA in every single voxel, and accept or reject the null hypothesis based on Lilliefors Test. Figure 17 shows, slice by slice, the voxels where the null hypothesis cannot be accepted, and they correspond to a

significant proportion of the slice voxel count (of the order of 40%—whereas, with $p = 0.05$, we only expect 5% to be deemed non-Gaussian by chance alone).

This is not so surprising given that we have tested the normality of the residuals on unsmoothed images. Rendering the data onto a normal distribution is one of the key motivations for smoothing the data. We thus repeat the Lilliefors test after smoothing the data with smoothing kernels ranging from 4 to 16 mm^3 FWHM. Smoothing does indeed reduce the number of voxels with non-Gaussian residuals, but the total number of

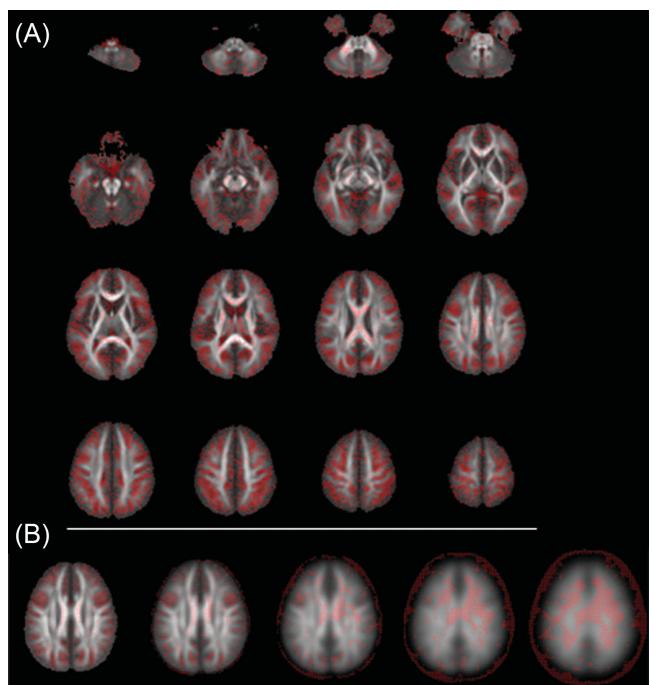


Figure 17. Red spots indicate the voxels where the Lilliefors test for the normality of residuals rejects the null hypothesis, i.e. where the residuals do not meet the assumption of normality. The large number of voxels with non normal residuals in un-smoothed images (A) does not decrease to the level expected due to chance (5% for $p = 0.05$) when smoothing the images with kernels ranging from 0–16 mm FWHM (B).

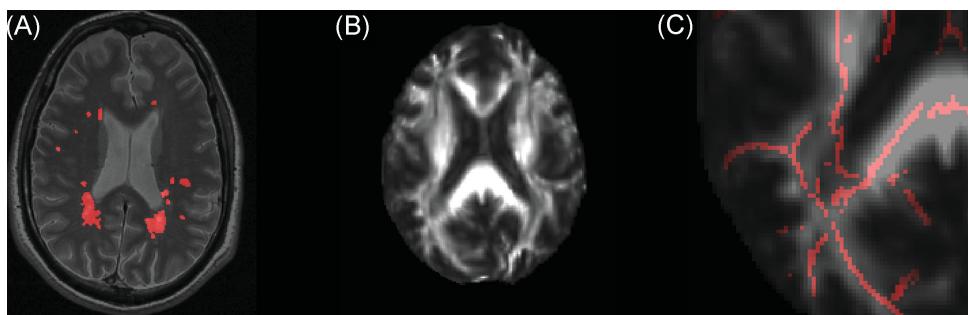


Figure 18. Effects of lesions on TBSS. Multiple white matter lesions are visible, shown in red in panel A. Under these circumstances, where FA tends to be reduced (B), it is unclear what values will be chosen as representative of the core of the tract during the process of FA projection in TBSS: Panel C shows a zoomed image of the FA skeleton superimposed onto the FA map in one of the areas affected by the disease.

voxels where the null hypothesis is rejected is still much higher than the 5% expected at $p = 0.05$.

PITFALL 23: DIFFERENCES IN REPORTING STATISTICS

Another issue to be considered is that the interpretation of the results strongly depends on the method used for correcting for multiple comparisons, and on the level of significance required. In order to reduce type I error due to the very large number of tests performed over all brain voxels, some correction for multiple comparison should be performed. However, there is no agreement on the most appropriate method. The most popular ones rely on the theory of Gaussian Random Fields (70), false discovery rate (71) and permutation-based approaches (72). Some researchers prefer to report only differences that reach significance at voxel level, others use more generous threshold that account also for the spatial distribution of supra-threshold voxels. Finally, it has become common practice to report uncorrected p values as well. These differing options yield a huge variety in the results obtained by different groups in similar experiments.

In summary, one can see that there are many pitfalls in applying the 'voxel-based morphometry' approach to analysis DT MRI data. At each stage in the processing pipeline (normalising, smoothing, statistical analysis), there are multiple options and the exact choice made at each step can have a profound influence on the outcome of the analysis. Put these multiple choices together

in a chain, and it becomes unsurprising that Jones *et al.* (73) found that when the same data set was sent out to nine different groups regularly using VB-type analyses, nine different results (each plausible) were obtained.

TBSS

In light of these issues, efforts have been made to make global search strategies more reliable. One such method is tract-based spatial statistics (TBSS) introduced by Smith *et al.* (74). TBSS has quickly become very popular, due to its availability (free with FSL, <http://www.fmrib.ox.ac.uk/fsl/>), the extremely simple input required from the user, and the ease of use (75). The novelty of TBSS consists of the introduction of an additional step, which follows image realignment into standard space. This step is the 'projection' of FA values from the main white matter tracts of all subjects onto an alignment-invariant tract representation, named the 'FA skeleton'. This projection is achieved by searching perpendicular to the local skeleton structure for the maximum value in the subject's FA image. This maximum value is assumed to represent the nearest relevant tract centre. This removes the need for smoothing the data, and increases the statistical power by reducing the total number of voxels to be tested. It should be noted that this step restricts the analysis to white matter only.

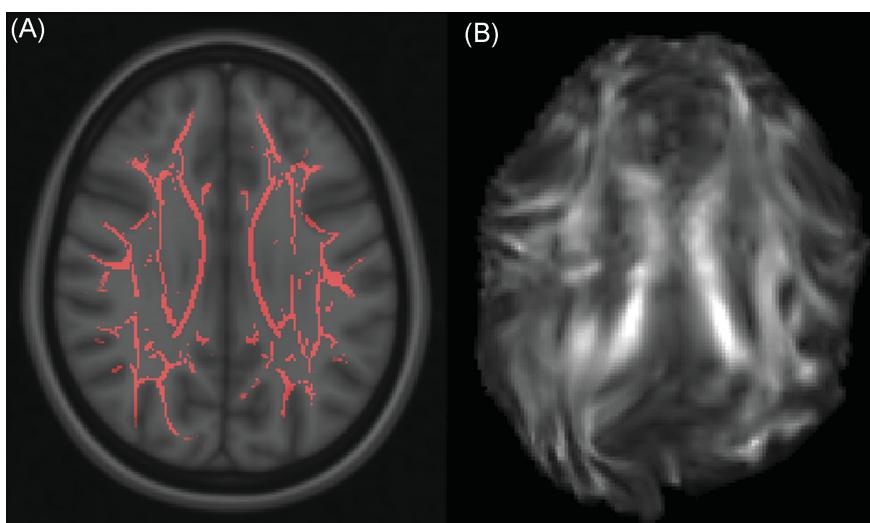


Figure 19. Results of TBSS. Inspecting the skeleton only (A) is not always enough to spot possible errors with registration. The skeleton in A is obtained from the normalised image shown in B.

PITFALL 24: IMPACT OF MACROSCOPIC LESIONS ON SKELETONISATION APPROACHES TO VB ANALYSIS

The process of skeletonisation relies on the identification of the maximum FA value in direction perpendicular to the direction of tract as derived from the average skeleton, followed by removal of low FA values (thresholding). If some white matter abnormalities are present (multiple sclerosis lesions, tumours, stroke), which are likely to locally reduce FA, their effects on the skeletonized FA image are difficult to predict. It is possible that some voxels which do not belong to the core of the tract have larger FA than those in the core because of the presence of the lesion (Fig. 18).

PITFALL 25: ERRORS IN SKELETONISATION

Finally, TBSS relies on the accurate coregistration of FA images, just like any other VB analysis method. Although it is tempting to trust the results of the pre-processing when no error message is given, and thus to review only the set of skeletonised images, produced almost completely automatically by the TBSS tools included in FSL, it should be noted that spotting registration errors in such images is almost impossible. The user is encouraged to verify the alignment of the maps before skeletonisation, as warping can go badly wrong when, for example, the quality of the input data is inadequate (Fig. 19).

CONCLUSION

This review article has assembled a number of reported and previously unreported potential pitfalls in the pipeline used for the analysis of diffusion MRI data. As we stated in our Introduction, this is not intended to be an exhaustive list [for example, we have intentionally not included many of the pitfalls in tractography analyses (the interested reader is referred to ref. (76) for more on this)], but is intended to provide wide coverage of many different aspects of the analysis pipeline. By highlighting so many pitfalls, the intention is not to convey the idea that analysis of diffusion MRI data is not robust and is always unreliable. To the contrary, in highlighting each pitfall, we have also indicated the source/contributing factor to the problem so that the reader can avoid it and obtain robust data and make sound interpretations. It is clear, however, that a researcher who is new to the field may obtain unexpected/surprising results from blindly treating a 'push-button' software package as a black-box, without understanding what lies beneath—or, worse, obtain dubious results without realising that they are dubious. If just one of these potential pitfalls is unknown to a reader, and they modify their analysis pipeline, or the way that they inspect their data along the way, then this article has served its purpose.

REFERENCES

- Le Bihan D, Breton E. *Imagerie de diffusion in vivo par résonance magnétique nucléaire*. C.R. Acad. Sc: Paris, 1985; T.301, Série II, 1109–1112.
- Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval Jeantet M. MR Imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*. 1986; 161: 401–407.
- Basser PJ, Mattiello J, Le Bihan D. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 1994; 66: 259–267.
- Basser PJ, Mattiello J, Le Bihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson.* 1994; 103: 247–254.
- Jones DK. Studying connections in the living human brain with diffusion MRI. *Cortex*. 2008; 44: 936–952.
- Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn. Reson. Med.* 1997; 38: 591–603.
- Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* 1999; 42: 952–962.
- Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. 135. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.* 2002; 47: 1202–1210.
- Reese TG, Heid O, Weisskoff RM, Wedeen VJ. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. *Magn. Reson. Med.* 2003; 49: 177–182.
- Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from B0 field variations. *Magn. Reson. Med.* 1995; 34: 65–73.
- Chang H, Fitzpatrick JM. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE Transact. Medic.Imaging*. 1992; 11: 319–329.
- Reber PJ, Wong EC, Buxton RB, Frank LR. Correction of off resonance-related distortion in echo-planar imaging using EPI-based field maps. *Magn. Reson. Med.* 1998; 39: 328–330.
- Morgan PS, Bowtell RW, McIntyre DJ, Worthington BS. Correction of spatial distortion in EPI due to inhomogeneous static magnetic fields using the reversed gradient method. *J. Magn. Reson. Imaging*. 2004; 19: 499–507.
- Embleton KV, Haroon HA, Morris DM, Ralph MA, Parker GJ. Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Hum Brain Mapp*. 2010 Feb 8.
- Bammer R, Auer M, Keeling SL, Augustin M, Stables LA, Prokesch RW, Stollberger R, Moseley ME, Fazekas F. Diffusion tensor imaging using single-shot SENSE-EPI. *Magn. Reson. Med.* 2002; 48: 128–36.
- Andersson JL, Skare S. A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. *NeuroImage*. 2002; 16: 177–199.
- Andersson JL, Richter M, Richter W, Skare S, Nunes RG, Robson MD, Behrens TE. Effects of susceptibility distortions on tractography. *Proceedings of the 12th Annual Meeting ISMRM*, Kyoto, Japan, 2004; 87.
- Lee J, Lazar M, Lee J, Holden J, Terasawa-Grilley E, Alexander AL. Correction of Bo EPI Distortions in Diffusion Tensor Imaging and White Matter Tractography. *Proceedings of the 12th Annual Meeting ISMRM*, Kyoto, Japan, 2004; 2172.
- Haselgrove JC, Moore JR. Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient. *Magn. Reson. Med.* 1996; 36: 960–964.
- Bastin ME. Correction of eddy current-induced artefacts in diffusion tensor imaging using iterative cross-correlation. *Magn. Reson. Imaging*. 1999; 17: 1011–1024.
- Rohde GK, Barnett AS, Basser PJ, Marenco S, Pierpaoli C. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Magn. Reson. Med.* 2004; 51: 103–114.
- Horsfield MA. Mapping eddy current induced fields for the correction of diffusion-weighted echo planar images. *Magn. Reson. Imaging*. 1999; 17: 1335–1345.
- Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med.* 1999; 42: 515–525.
- Jones DK. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: A Monte Carlo study. *Magn. Reson. Med.* 2004; 51: 807–815.
- Leemans A, Jones DK. The B-matrix must be rotated when motion correcting diffusion tensor imaging data. *Magn. Reson. Med.* 2009; 61: 1336–1349.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edition. Cambridge University Press: New York, 1994; 183.
- Davis TL, Wedeen VJ, Weisskoff Rosen BR. White matter tract visualization by echo-planar MRI. In *Book of Abstracts: Twelfth Annual Meeting of the Society of Magnetic Resonance in Medicine*. ISMRM: Berkeley, CA, 1993; p. 289.
- Bevington PR, Robinson DK. *Data Reduction and Error Analysis for the Physical Sciences*, 2nd edition. McGraw-Hill: New York, 1992.

29. Papadakis NG, Xing D, Houston GC, Smith JM, Smith MI, James MF, Parsons AA, Huang CL, Hall LD, Carpenter TA. A study of rotationally invariant and symmetric indices of diffusion anisotropy. *Magn. Reson. Imaging.* 1999; 17: 881–892.
30. Hasan KM, Alexander AL, Narayana PA. Does fractional anisotropy have better noise immunity characteristics than relative anisotropy in diffusion tensor MRI? An analytical approach. *Magn. Reson. Med.* 2004; 51: 413–417.
31. Shimony JS, McKinstry RC, Akbudak E, Aronovitz JA, Snyder AZ, Lori NF, Cull TS, Conturo TE. Quantitative diffusion-tensor anisotropy brain MR imaging: normative human data and anatomic analysis. *Radiology.* 1999; 212: 770–784.
32. Lee CE, Danielian LE, Thomasson D, Baker EH. Normal regional fractional anisotropy and apparent diffusion coefficient of the brain measured on a 3 T MR scanner. *Neuroradiol.* 2009; 51: 3–9.
33. Yasmin H, Aoki S, Abe O, Nakata Y, Hayashi N, Masutani Y, Goto M, Ohtomo K. Tract-specific analysis of white matter pathways in healthy subjects: a pilot study using diffusion tensor MRI. *Neuroradiol.* 2009; 51: 831–840.
34. Pierpaoli C, Jezzard P, Basser PJ, Barnett AS. Diffusion tensor MR imaging of the human brain. *Radiology.* 1996; 201: 637–648.
35. Yoshiura T, Wu O, Zaheer A, Rees TG, Sorenson AG. Highly diffusion-sensitized MRI of brain: dissociation of gray and white matter. *Magn. Reson. Med.* 2001; 45: 734–740.
36. Pierpaoli C, Basser PJ. Towards a quantitative assessment of diffusion anisotropy. *Magn. Reson. Med.* 1996; 36: 893–906.
37. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn. Reson. Med.* 1995; 34: 910–914.
38. Dietrich O, Heiland S, Sartor K. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. *Magn. Reson. Med.* 2001; 45: 448–453.
39. Jones DK, Basser PJ. ‘Squashing peanuts and smashing pumpkins’: how noise distorts diffusion-weighted MR data. *Magn. Reson. Med.* 2004; 52: 979–993.
40. Jones DK. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magn. Reson. Med.* 2003; 49: 7–12.
41. Schwartzman A, Dougherty RF, Taylor JE. Cross-subject comparison of principal diffusion direction maps. *Magn. Reson. Med.* 2005; 53: 1423–1431.
42. Whitcher B, Wisco JJ, Hadjikhani N, Tuch DS. Statistical group comparison of diffusion tensors via multivariate hypothesis testing. *Magn. Reson. Med.* 2007; 57: 1065–1074.
43. Rashid W, Hadjipropis A, Griffin CM, Chard DT, Davies GR, Barker GJ, Tofts PS, Thompson AJ, Miller DH. Diffusion tensor imaging of early relapsing-remitting multiple sclerosis with histogram analysis using automated segmentation and brain volume correction. *Mult. Scler.* 2004; 10: 9–15.
44. Jones DK, Catani M, Pierpaoli C, Reeves SJC, Shergill SS, O’Sullivan MO, Malley JD, McGuire P, Horsfield MA, Simmons A, Williams SCR, Howard RJ. 2005. Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. *Hum. Brain Mapp.* 2005; 273: 230–238.
45. Shergill SS, Kanaan RAA, Chitnis XA, O’Daly O, Jones DK, Frangou S, Williams SCR, Howard RJ, Barker GJ, Murray RM, McGuire P. A diffusion tensor imaging study of fasciculi in schizophrenia. *Am. J. Psychiatry.* 2007; 164: 467–473.
46. Kanaan RA, Shergill SS, Barker GJ, Catani M, Ng VW, Howard RJ, McGuire PK, Jones DK. Tract-specific anisotropy measurements in diffusion tensor imaging. *Psychiatry Res. Neuroimaging.* 2006; 146: 73–82.
47. Jones DK, Travis AR, Eden GM, Pierpaoli C, Basser PJ. PASTA: Pointwise assessment of streamline tractography attributes. *Magn. Reson. Med.* 2005; 53: 1462–1467.
48. Liu G, van Gelderen P, Duyn J, Moonen CTW. Single-shot diffusion MRI of human brain on a conventional clinical instrument. *Magn. Reson. Med.* 1996; 35: 671–677.
49. Bhagat YA, Beaulieu C. Diffusion anisotropy in subcortical white matter and cortical gray matter: changes with aging and the role of CSF-suppression. *J. Magn. Reson. Imaging.* 2004; 20: 216–227.
50. Steens SC, Admiraal-Behloul F, Schaap JA, Hoogenraad FG, Wheeler-Kingshott CA, le Cessie S, Tofts PS, van Buchem MA. Reproducibility of brain ADC histograms. *Eur. Radiol.* 2004; 14: 425–430.
51. Concha L, Gross DW, Beaulieu C. Diffusion tensor tractography of the limbic system. *AJR Am. J. Neuroradiol.* 2005; 26: 2267–2274.
52. Chou MC, Lin YR, Huang TY, Wang CY, Chung HW, Juan CJ, Chen CY. FLAIR diffusion-tensor MR tractography: comparison of fiber tracking with conventional imaging. *AJR Am. J. Neuroradiol.* 2005; 26: 591–597.
53. Malykhin N, Concha L, Seres P, Beaulieu C, Coupland NJ. Diffusion tensor imaging tractography and reliability analysis for limbic and paralimbic white matter tracts. *Psychiatry Res.* 2008; 164: 132–142.
54. Pierpaoli C, Jones DK. Removing CSF Contamination in Brain DT-MRIs by Using a Two-Compartment Tensor Model. In *Proceedings of the 12th Annual Meeting ISMRM, Kyoto*, 2004; 1215.
55. Pasternak O, Sochen N, Gur Y, Intrator N, Assaf Y. Free water elimination and mapping from diffusion MRI. *Magn. Reson. Med.* 2009; 62: 717–730.
56. Ashburner J, Friston KJ. Voxel-based morphometry—the methods. *NeuroImage.* 2000; 11: 805–821.
57. Bookstein FL. ‘Voxel-based morphometry’ should not be used with imperfectly registered images. *NeuroImage.* 2001; 14: 1454–1462.
58. Ashburner J, Friston KJ. Why voxel-based morphometry should be used. *NeuroImage.* 2001; 14: 1238–1243.
59. Davatzikos C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage.* 2004; 23: 17–20.
60. Park HJ, Kubicki M, Shenton ME, Guimond A, McCarley RW, Maier SE, Kikinis R, Jolesz FA, Westin CF. Spatial normalization of diffusion tensor MRI using multiple channels. *NeuroImage.* 2003; 20: 1995–2009.
61. Guimond A, Guttmann CRG. Deformable registration of DT-MRI data based on transformation invariant tensor characteristics. *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI’02), Washington, DC, USA.* 2002.
62. Alexander DC, Pierpaoli C, Basser PJ, Gee JC. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Trans. Med. Imaging.* 2001; 20: 1131–1139.
63. Kawashima T, Nakamura M, Bouix S, Kubicki M, Salisbury DF, Westin CF, McCarley RW, Shenton ME. Uncinate fasciculus abnormalities in recent onset schizophrenia and affective psychosis: a diffusion tensor imaging study. *Schizophr. Res.* 2009; 110: 119–126.
64. Fujiwara H, Namiki C, Hirao K, Miyata J, Shimizu M, Fukuyama H, Sawamoto N, Hayashi T, Murai T. Anterior and posterior cingulum abnormalities and their association with psychopathology in schizophrenia: a diffusion tensor imaging study. *Schizophr. Res.* 2007; 95: 215–222.
65. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* 1996; 4: 58–73.
66. Friston KJ, Holmes AP, Worsley KJ, Poline J-B, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 1995; 2: 189–210.
67. Rosenfeld A, Kak AC. *Digital Picture Processing 2*. Academic Press: Orlando, FL, 1982; 42.
68. Jones DK, Symms MR, Cercignani M, Howard RJ. 2005. The effect of filter size on the outcome of VBM analyses of DT-MRI data. *NeuroImage.* 2005; 26: 546–554.
69. Lilliefors H. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Statistic. Assoc.* 1967; 62: 399–402.
70. Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* 1992; 12: 900–918.
71. Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage.* 2002; 15: 870–878.
72. Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 2002; 15: 1–25.
73. Jones DK, Chitnis XA, Job D, Khong PL, Leung LT, Marenco S, Smith SM, Symms MR. (what happens when nine different groups analyze the same DT-MRI data set using voxel-based methods?). In *Proceedings of the 15th Annual Meeting ISMRM, Berlin*, 2007; 74.
74. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM and others. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage.* 2006; 31: 1487–1450.
75. Smith SM, Johansen-Berg H, Jenkinson M, Rueckert D, Nichols TE, Miller KL, Robson MD, Jones DK, Klein JC, Bartsch AJ, Behrens TEJ.