What Happens When Nine Different Groups Analyze the Same DT-MRI Data Set Using Voxel-Based Methods?

D. K. Jones[1], X. A. Chitnis[2], D. Job[3], P. L. Khong[4], L. T. Leung[5], S. Marenco[6], S. M. Smith[7], and M. R. Symms[8]

[1]Cardiff University Brain Repair and Imaging Centre, School of Psychology, Cardiff, Wales, United Kingdom, [2]Centre for Neuroimaging Sciences, Institute of Psychiatry, London, United Kingdom, [3]Division of Psychiatry, Edinburgh University, Edinburgh, United Kingdom, [4]Department of Diagnostic Radiology, University of Hong Kong, Hong Kong, [5]Department of Diagnostic Radiology, Univeristy of Hong Kong, Hong Kong, [6]GCAP, Clinical Brain Disorders Branch, National Institute of Mental Health, Bethesda, MD, United States, [7]fMRIB Centre, Oxford University, Oxford, United Kingdom, [8]Institute of Neurology, University College London, London, United Kingdom

INTRODUCTION: There are two main ways to perform group comparisons of DT-MRI data, i.e. (1) local region of interest (ROI)-based analyses; and (2) whole brain voxel-based global search approaches. The former are appropriate when the prior information is sufficiently robust to allow prediction of the location and extent of expected differences. However, in many cases (particularly psychiatric disorders) the spatial location / extent of differences are unknown *a priori*. Consequently, whole brain voxel-based analyses are performed, comparing groups on a voxel-by-voxel basis – essentially checking every brain location for patient/control differences and obviating the need for any *a priori* hypotheses. However, the user of such an approach is presented with a large choice of parameters in setting up the comparisons, including (but not limited to): the form of spatial normalization (e.g., affine *vs.* nonlinear deformations); the template used for spatial normalization (e.g. an EPI-based template or a high resolution structural scan); masking (is the search volume confined to a particular tissue type?), smoothing (what degree of spatial filtering is used), statistical testing (parametric or non-parametric approaches?); and the metric used to report the result. This huge choice leads to considerable heterogeneity of methods in the diffusion literature. Moreover, there is also heterogeneity in the diffusion literature as to what DT-MRI tells us about group differences in particular diseases/ conditions, one of the most marked examples being in schizophrenia where a huge variety of results have been found with DT-MRI. It is unclear whether this heterogeneity results from (a) heterogeneity in the population samples across different studies; or (b) heterogeneity in methods used to look for group differences, or both.

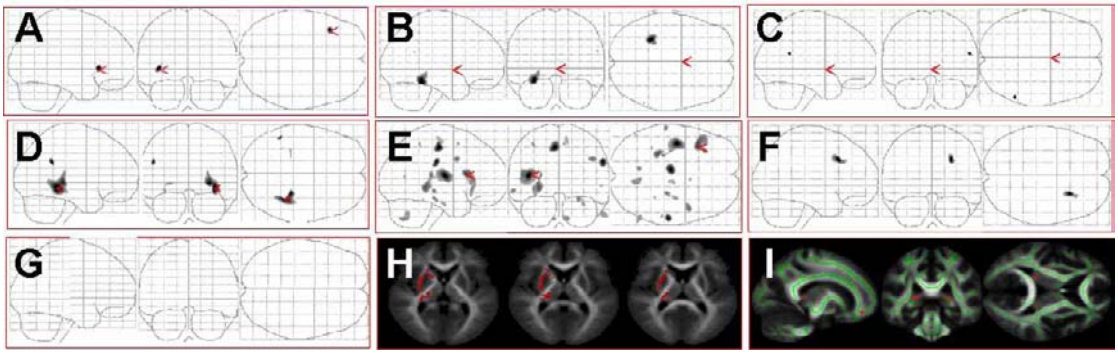
The question asked in the present study is very simple: What if we remove the effects of heterogeneity in the patient /control sample by asking different groups to analyse the same data set using their preferred voxel-based approach? Do the different groups find the same results – or does the choice of design give different results?

METHOD: Subjects: We recruited 14 healthy right handed schizophrenic (DSM-IV) males (mean age 34 years, range: 22-53 years); median IQ =110 (range 98-124) and 14 healthy right hand male control subjects (mean age 34 years, range: 19-57 years; median IQ = 109, range 99-123). Data Acquisition: Whole brain isotropic (2.5 mm^3) DT-MRI data were acquired from each subject using a GE Signa 1.5T system and a gated optimised DT-MRI sequence[1]. Following distortion correction, the diffusion tensor was determined in each voxel[2] and images of fractional anisotropy[3] computed for each subject. Analysis: The data sets were sent blind to nine different labs for analysis, the only guidance being that the data were collected from 2 different groups – and that each lab should use their 'usual voxel-based approach' to compare the 2 groups. The salient features of the analysis methods are presented opposite:

Method	Package	Normalization	Smooth	Threshold
A	SPM99[4]	Affine	12 mm	p= 0.05 (corr)
B	SPM99[4]	Nonlinear	12 mm	p= 0.05 (corr)
C	SPM99[4]	Nonlinear	7 mm	p= 0.05 (corr)
D	SPM2[4]	Nonlinear	12 mm	p= 0.05 (FDR)
E	SPM99[4]	Nonlinear	12 mm	p= 0.05 (uncorr)
F	SnPM2[5]	Nonlinear	6 mm	p= 0.05(corr)
G	FSL[6]	Nonlinear	7 mm	p= 0.05 (corr)
H	SBAM[7]	Nonlinear	4 mm	p= 0.05 (corr)
I	TBSS[8]	Nonlinear+Projection	None	p= 0.05 (corr)

RESULTS: The results obtained using 9 the different methods are presented in Figure 1. In summary, Method A found significant reductions in the vicinity of the **left arcuate fasciculus**. Method B found FA reductions predominantly in left **cerebellum /fusiform gyrus**. Method C found reductions in **right supramarginal / angular gyrus**. Method D found reduced FA predominantly in the **right lateral occipito-temporal junction**. Method E found widespread group differences prior to correction for multiple comparisons. As this laboratory routinely used the small volume correction (SVC), they were then informed that one of the subject groups was of schizophrenics. They then applied their SVC approach centred on the region of **the left arcuate fasciculus –** and found a significant effect. Method F found reduced FA in the vicinity of the right **superior longitudinal fasciculus**. However, these data were then reanalyzed with exactly the same method after co-varying for the whole brain mean FA, and then **no significant group effects** remained. Method G used the standard t-test approach included in the FSL package – and found **no group differences**. Method H found reduced FA in **the internal/external capsule**. Finally, Method I – using TBSS[8] found reduced FA in the **posterior pillar of the fornix bilaterally, left posterior external/extremal capsule, and left inferior fronto-occipital fasciculus / corpus callosum.**

FIGURE 1: Areas of reduced anisotropy in schizophrenics. The green voxels in I highlight the skeletonised search volume, while the orange voxels indicate reduced anisotropy in schizophrenics.



DISCUSSION: There is *minimal* overlap between results obtained with the 9 different methods. This reinforces the oft-overlooked heterogeneity of results in the DT-MRI literature introduced simply by the parameters chosen/ method used for voxel-based analyses. The majority of the positive results here could be consistent with a schizophrenia hypothesis, and each user carefully chose their analysis-parameters. This serves as a reminder of what is being tested under the null hypothesis, i.e. just because one method finds a particular difference, it does NOT mean that there were NO other differences – a fact that can be easily overlooked. This study therefore also highlights the extreme difficulty in interpreting differences in reported results obtained by different labs where there may be differences in the subjects recruited, or in the methods chosen to analyse the data, or indeed both. We hope this therefore has served as a useful exercise to those planning a voxel-based study of DT-MRI data in the future.

REFERENCES: [1] Jones et al. *Hum Brain Mapp* 2002, **15:**216- ; [2] Basser et al. *Biophys J.* 1994, **66:**259- ; [3] Basser et al. J Magn Reson B 1996, **111:**209- ; [4] Friston et al. Hum Brain Mapp 1995, **2:**189-; [5] Holmes et al. *JCBF* 1996, **16:**7-; [6] Smith et al. *NeuroImage* 2004; **23:**S208**-;** [7] Bullmore et al. *TMI* 1999;**18:**32-; [8] Smith et al. *NeuroImage* 2006; **31:**1487- ;