Author's response to reviews

Title: Distribution of tract deficits in schizophrenia

Authors:

Ian Ellison-Wright (ian.ew@inbox.com)
Pradeep J Nathan (pn254@cam.ac.uk)
Edward T Bullmore (etb23@cam.ac.uk)
Rashid Zaman (rz218@cam.ac.uk)
Robert Dudas (rbd21@cam.ac.uk)
Mark Agius (ma393@cam.ac.uk)
Emilio Fernandez-Egea (Emilio.Fernandez@cpft.nhs.uk)
Ulrich Müller (um207@cam.ac.uk)
Chris M Dodds (Chris.M.Dodds@gsk.com)
Natalie J Forde (plugkicker@gmail.com)
Cathy Scanlon (cathy.scanlon@nuigalway.ie)
Alexander Leemans (Alexander@isi.uu.nl)
Colm McDonald (colm.mcdonald@nuigalway.ie)
Dara M Cannon (dara.cannon@nuigalway.ie)

Version: 3 Date: 13 February 2014

Author's response to reviews: see over
The Editor
BMC Psychiatry
13 February 2014

Dear Sir,

Re: MS: 1475091454112466  Distribution of tract deficits in schizophrenia

Thank you for your recent letter regarding our manuscript. We thank the reviewers for their insightful comments, and in summary are in full agreement with their recommendations.

We have responded to each individual point below and made the appropriate changes to the manuscript (highlighted).

Reviewer: Richard Antony Alexander Kanaan

Major Compulsory Revisions

1. Methods – a power analysis is clearly missing. It’s not essential that the study is properly powered (it probably isn’t) but the results are very hard to interpret without it.

We have now included a power analysis:

‘Power analysis, based on fractional anisotropy values detected in a previous large study [28], indicated that a sample of 23 subjects in each group would provide statistical power over 80% (testing for reduced FA in the corpus callosum in schizophrenia; one-tailed test, alpha level 5%). From the original sample, 21 patients and 21 controls completed the imaging protocol (reducing power to 70%).’

2. Methods – were the volumetric scans that were excluded for motion artefact still employed for the tractography definition process? How can this be justified?

The excluded volumetric scans were not employed for tractography measurements (and were also excluded, as described, from the VBM analyses) although they were of adequate quality for defining the anatomical gates.

The separate diffusion images were corrected for subject motion and eddy current induced distortions and these images were of a quality eligible for tractography measurements.

3. Results – there appears to be a huge multiple comparisons problem with the regional analysis, and I suspect any reasonable approach to correcting this would eliminate all results.

We have clarified that (i) the voxel-based analyses were the primary analyses and (ii) the statistical values provided for the tractography and regional distribution of FA values were indicative rather than for strict hypothesis testing.
We have revised the Methods section to read:

‘Our primary analyses were of: (i) the voxel-based analysis of gray matter between the two groups, (ii) the voxel-based analysis of fractional anisotropy and (iii) the tract FA in the right and left ALIC. Our secondary analyses were of the FA values in the corpus callosum, splenium, SLF and cingulum bundle. Our tertiary analyses were of the regional distribution of FA along the TBSS skeleton in each region of the White Matter Atlas. As the study employed multiple analysis methods and many brain volume measurements are correlated, we did not correct for multiple testing (other than the False Discovery Rate used in the voxel-based analyses) and the probability statistics need to be interpreted in this context.’

4. Results – was the tract data normally distributed – the authors mention testing this, but don’t give the results – to justify using an ANCOVA?

The median FA for each tract (analysed by ANCOVA) was normally distributed.

In response to this query, we have revised the Methods section to read:
‘Fractional anisotropy is non-parametrically distributed across tracts so the median FA was employed for each tract in each person rather than mean.

The Shapiro-Wilks test and Levene's test were used to determine normality of distribution and homogeneity, respectively, for median FA of the tracts and age of subjects. An analysis of co-variance (ANCOVA) was used to compare diagnostic groups, covarying for age and tract volume. Independent samples t-test was used to test for differences in tract FA between diagnostic groups. Mann Whitney U was employed to compare age between groups. A chi-square test was used to assess the gender proportions between diagnostic groups.’

We have added the following to the results section:
‘(age was non-normally distributed, W=0.94, p=0.02),
‘Median FA for each tract examined was normally distributed (W=0.97-0.98, p=0.26-0.61).’

Discretionary Revisions

5. Background – this is under-referenced. The authors should acknowledge the wide range of tracts implicated in previous studies – not just the ones they are interested in.

We have now described the wide range of tracts implicated in previous studies bringing these up to date with references to major review papers and the latest research studies in the field.

6. Background – why discuss spatial progression when this study does not appear relevant to its investigation?

For greater clarity we have now described this in the Future Research part of the Discussion section and removed reference to this in the Abstract Background and Conclusion.

We are interested in white matter tract changes in relation to spatial progression because, firstly, we have observed that there are similarities between the spatial pattern of gray matter changes in schizophrenia and fronto-temporal dementia and, secondly, studies of neurodegenerative disorders have suggested that the spatial pattern of disease progression is related to the connectivity of large-scale white matter networks. We have rewritten this section in the Discussion to explain this more clearly.

7. Background – was there ever an intention to combine the 3 methods? It’s not clear how these were meant to complement each other. Why use the tractography at all, when the regional method you have gives meaningful localisation?
We designed the analysis to utilise the three methods. We used TBSS as an efficient brain-wide and spatially neutral method to detect clusters of change in the patients with schizophrenia. The tractography was a labour-intensive but anatomically-guided method used to detect tract-wide deficits which may be missed where the skeleton under-represents parts of tracts and secondary tracts or where the TBSS underestimates changes at locations where several tracts cross. The regional analysis provided data for comparison with other studies (e.g. by meta-analysis) to allow quantification of the FA changes in each white matter region.

We have rewritten this section of the Introduction:
‘Our objective was to combine a brain-wide voxel-based approach (TBSS) (which is highly automated) to guide subsequent tract-based analysis (which is more labour-intensive but has the potential for greater sensitivity in detecting changes because of greater anatomical specificity) as well as an atlas-based analysis (permitting quantification of the magnitude of regional changes and providing data for future meta-analyses).’

8. Methods – why include subjects with schizoaffective disorder? Though the effects of two subjects may be small, they should really consider some kind of sensitivity analysis.

Although an ideal sample would include just subjects with schizophrenia, including subjects with related diagnoses increases the power of the sample. We followed the precedent of the Voxel-based Morphometric Multisite Collaborative Study on Schizophrenia which included some subjects with schizoaffective disorder as studies have shown that those with schizoaffective disorder have functional outcome similar to schizophrenia, there is a significant overlap between the diagnoses, and diagnostic differentiation can be difficult.

We have now conducted a sensitivity analysis of the TBSS results by removing the two subjects with schizoaffective disorder. This did not change either the direction or the spatial location of the peak statistical differences detected between groups. We have included this in the Results section.

9. Methods – the statement that additional medications were an exclusion unless the investigator didn’t think so seems completely meaningless to me. Either clarify, or drop it.

For clarity we have removed this sentence.

10. Methods – I don’t understand the nature of the tractography process used, and its talk of ‘gates’. The authors should provide some broad outline of how the software worked for those unfamiliar with the method.

We have included a more detailed explanation of this in the Methods section:
‘The tracts were defined anatomically as detailed below using ‘inclusion gates’. For each tract in each individual subject, two anatomical structures were defined in the brain (‘gates’) based on defined anatomical landmarks. A tract consisted of all reconstructed paths passing through both anatomically defined gates.’

11. Methods – an intra-rater reliability is mentioned – what was this defined on (which tractography measure)?

A single rater (NF) defined all tracts. Her reliability and consistency in defining these anatomically was examined by blinding the rater to the images and re-defining the same tracts twice - this achieved an acceptable level of reliability in the independence and consistency of tracing as evidenced by the intraclass correlation coefficient of at least 86% on the worst and 98% on the best cases across different tracts.

For clarity we have given a more detailed explanation in the Methods section:
‘A single rater (NF) defined all tracts. Her reliability and consistency in defining these anatomically was examined by blinding the rater to the images and re-defining the same tracts twice - this achieved an acceptable level of reliability of greater than 86% in defining tracts (ICC= 0.86-0.99).’

12. Methods - Why was FDR used for the VBM analysis, and not SPM’s more usual FWE?

In both the grey matter and the FA white matter voxel-based analyses FDR based correction was employed. Correcting for the rate of false discoveries and detecting family wise errors are both widely accepted methods used to correct for multiple comparisons in the neuroimaging literature (Whitwell, 2009). We have added this reference to the Methods section.

The advantages of the two approaches have been debated by the SPM-VBM community. The FWE correction controls the chance of any false positives across the entire volume, but may be over-stringent, whereas the FDR correction controls the expected proportion of false positives among suprathreshold voxels.

13. Methods – it is unclear exactly which data was taken from the TBSS through to the regional analysis.

The tracts selected, defined and analysed were based on them containing the spatial location of the peak T-statistic values in the output of the voxel-based analysis of FA maps.

14. Methods – why was age used as a covariate for the VBA but not for its regional comparison?

Age was included as a covariate for both the voxel based analysis of FA and the tract-based ANCOVA.

We have clarified this in the Methods section.

Reviewer: Julia Friederike Sowislo

First, I ask the authors to elaborate more on the novelty of the findings (in contrast to the novelty of the method).

We have now addressed this in point 1 below.

Second, it seems to me that the article addresses a rather specialized audience. As can been seen from my below comments, I would suggest that the author clarify some issues in order to make the article readable for a wider audience. I think that clarifications would strengthen the contribution of the current research.

We thank the reviewer for this observation and have extensively rewritten the Introduction and have addressed the issues below in order to make the paper more accessible to the general reader.

Minor Essential Revisions

(1) I would like the authors to precisely indicate what can be learnt from this article over and above what is known from previous meta-analysis of DTI in schizophrenia (e.g., Ellison- Wright & Bullmore, 2009; Kanaan et al., 2005).

We have now addressed the novelty of the findings of this study more directly in our Discussion:
Compared with previous studies of FA changes in schizophrenia, this study found changes in medial frontal regions but did not identify FA reductions in other regions which have been previously implicated. As the sample size was small, this may indicate that the medial frontal regions are the areas of maximal FA change in schizophrenia.

Another notable aspect of the results was that we did not find gray matter reductions in patients compared with controls, unlike many previous studies of gray matter changes in schizophrenia. As discussed below, this may be attributable to the high proportion of patients treated with atypical antipsychotics. It suggests that in these patients, subtle diffusion changes may be easier to detect than volumetric changes.

2) The authors claim that one major purpose of the article was to relate white matter tract differences in schizophrenia to changes in gray matter (p.5). However, as far as I understood, the study did not detect any gray matter changes in schizophrenics as compared to controls. Can the authors discuss what the null-finding in gray matter changes means with regard to the initial research purpose?

The lack of voxel-based gray matter findings was unexpected as most VBM studies have detected changes. As noted in the paper, all patients were receiving atypical antipsychotics. Our sample of patients may also have been unusual because of the large number of patients taking clozapine (8 of 23) – which we have now added in the paper:

‘Antipsychotic treatment has been associated with complex regional gray matter changes (both increases and decreases) [81]; however, gray matter loss may be less intense and widespread in patients treated with atypical antipsychotics (olanzapine) compared to typical antipsychotics (haloperidol) [82]. 8 of 21 patients in our sample were prescribed clozapine and clozapine treatment has been associated both with gray matter increases [83] or with less gray matter loss over time [84]. A study of first-episode psychosis subjects treated with atypical antipsychotics also found FA decreases but no gray matter changes [80] whereas studies of medication-naïve patients with schizophrenia have found evidence of gray matter deficits [44,85]. It is also of note that the mean premorbid IQ of the patients was high; higher performance IQ has been positively correlated with FA in schizophrenia [86].’

3) Relatedly, as the study could not replicate the previously found volumetric changes in gray and white matter, I wonder whether the study might be underpowered. The authors discussed the drop-out of seven subjects and the possibility of type 2 error. Is there any evidence for the drop-out being selective?

For the structural VBM gray and white matter analyses, 7 images were excluded. These were from 5 of 21 patients and 2 of 21 controls. This reduced the power of the VBM analysis and may have contributed to the lack of positive findings.

Could the authors give justification for the chosen sample size (i.e., show the results of a power analysis)?

This is discussed in point 1 (Reviewer 1): the sample size was designed to provide statistical power of over 80% (although not all subjects completed the imaging protocol, reducing power to 70%).

4) White matter changes were found using diffusion analyses but not using volumetry. Can the authors give some examples for less familiar readers demonstrating why changes in fractional anisotropy do not necessarily have to result in volumetric changes?

We have provided more information about this in the Discussion:

‘White matter FA and volumetric changes in schizophrenia may not necessarily be correlated [75]. Abnormal FA may reflect changes in integrity of the myelin sheath and axonal membrane but, in general, there are many factors that can modulate the FA [26]. In neurodegenerative disorders, while
FA changes often correlate with atrophy they may also be found without volumetric changes depending on the methodology, region studied and underlying pathology [76]. For example, in Alzheimer’s disease, some regions of white matter FA decrease may reflect microstructural changes rather than macroscopic changes [77] and this may be a marker for future atrophy [78]. A meta-analysis comparing white matter volumetry and DTI changes in schizophrenia suggested that DTI studies appeared more sensitive to white matter abnormalities in schizophrenia [5].

(5) In addition to the 21 patients who met diagnosis of schizophrenia, the authors included two patients who met diagnosis of schizoaffective disorder. I would question the authors to explain why they included these two schizoaffective patients.

This is discussed in point 8 (Reviewer 1), we have now conducted a sensitivity analysis by removing the two subjects with schizoaffective disorder. This did not change the direction or spatial location of the peak statistical differences detected between groups.

6) The authors state that the control participants were matched (p.3). They report that there were no significant differences in variables that are supposed to affect fractional anisotropy such as age (e.g., Pefferbaum et al., 2000), IQ (e.g., Schmithorst et al., 2005), gender, and handedness (e.g., Westerhausen et al., 2003). However, the groups do not seem to be matched with regard to these variables (e.g., 17 males in the study group vs. 14 males in the study group).

Can the authors indicate how the matching was done?

We thank the reviewer for drawing our attention to these references which we have included.

We aimed to match the groups for age, gender and premorbid IQ. The groups did not differ significantly in mean age (p=0.39) or IQ (p=0.25). All subjects were right-handed (Edinburgh Handedness Inventory).

Although there were more males in the study group, the groups did not differ significantly in gender proportion (p=0.30). As male gender is associated with higher FA (den Baber et al, 2013), this should not have confounded our result (of lower FA in patients) but may have reduced our power to detect a difference. We have now referred to this in our Discussion (see point 8 below).

(7) The authors write that “either independent sample t-test or Mann Whitney U was employed to compare age between groups” (p.14). If I understood correctly, they tested for age differences between the two study groups. I would like the authors to indicate which of both tests they employed.

We agree that our explanation was ambiguous. Independent samples t-test was used where normality of distribution was determined and the non-parametric equivalent was used where the distribution of the variable was not normal, that is, Mann-Whitney U.

As discussed in point 4 (Reviewer 1), we have clarified our explanation of the statistical tests.

Discretionary Revisions

(8) The authors argue that some structural changes in schizophrenia may not represent a primary pathology, but result from the symptoms. To shed more light on these issues the authors suggest conducting studies that examine potential confounders. I would ask the authors to give some examples of confounding variables.

Would it be possible to add some additional analyses to the study which include some of these variables?

We have provided some examples of confounding variables with additional analyses:
‘For example, if increased psychotic symptoms cause FA decreases, then there may be a correlation between symptomatology and brain structural measures. We conducted an exploratory test on our data by testing for a correlation between corpus callosum genu median FA and PANSS score (a measure of psychotic symptoms). We found a weak negative correlation which was not significant (Pearson’s bivariate correlation -0.26, p=0.08). These effects may be further elucidated by examining potential confounders or rescanning subjects during periods of exacerbation and remission of symptoms.

There were a number of other potential methodological limitations. The use of antipsychotic treatment in the patients may be a confounding factor so diffusion changes could potentially be due to the effect of medication rather than diagnosis. FA decreases have been found in the anterior cingulate and right corona radiata in previously drug-naive patients with schizophrenia after six weeks of treatment [104]. Other potential confounders are age [105-107], IQ [108], gender [109], and handedness [110]. In our data, examining the corpus callosum genu median FA as a measure of structural change, age was negatively correlated, (Pearson’s bivariate correlation -0.38, p=0.01, whole group) and IQ was positively correlated (Pearson’s bivariate correlation 0.31, p=0.05, whole group) as expected. The patient and control groups in our study were well matched for age and IQ so this should not have affected our results. In this study, corpus callosum genu median FA did not differ significantly by gender (t-test t(41)-0.86, p=0.39) – the average FA was 2% greater in males than females (whole group). The slightly greater proportion of males in the schizophrenia group would be expected to result in higher FA values [109], rather than reduced FA values (which we detected) so this slight gender imbalance may have reduced our power to detect changes.’

Thank you again for considering this paper for publication in BMC Psychiatry.

With kindest regards,

Dr Ian Ellison-Wright
On behalf of all authors.