Voxel-based morphometry (VBM) studies in schizophrenia—can white matter changes be reliably detected with VBM?

Eric D. Melonakos a, Martha E. Shenton a,b, Yogesh Rathi a, Douglas P. Terry a, Sylvain Bouix a, Marek Kubicki a,b,*

a Psychiatry Neuroimaging Laboratory, Department of Psychiatry, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA
b Department of Psychiatry, VA Boston Healthcare System, Harvard Medical School, Boston, MA, USA

ARTICLE INFO
Article history:
Received 27 July 2010
Received in revised form 14 January 2011
Accepted 14 January 2011

Keywords:
Diffusion tensor imaging
Meta-analysis
Chronic schizophrenia
First-episode schizophrenia
Fractional anisotropy

ABSTRACT
Voxel-based morphometry (VBM) is a hypothesis-free, whole-brain, voxel-by-voxel analytic method that attempts to compare imaging data between populations. Schizophrenia studies have utilized this method to localize differences in diffusion tensor imaging (DTI) derived fractional anisotropy (FA), a measure of white matter integrity, between patients and healthy controls. The number of publications has grown, although it is unclear how reliable and reproducible this method is, given the subtle white matter abnormalities expected in schizophrenia. Here we analyze and combine results from 23 studies published to date that use VBM to study schizophrenia in order to evaluate the reproducibility of this method in DTI analysis. Coordinates of each region reported in DTI VBM studies published thus far in schizophrenia were plotted onto a Montreal Neurological Institute atlas, and their anatomical locations were recorded. Results indicated that the reductions of FA in patients with schizophrenia were scattered across the brain. Moreover, even the most consistently reported regions were reported independently in less than 35% of the articles studied. Other instances of reduced FA were replicated at an even lower rate. Our findings demonstrate striking inconsistency, with none of the regions reported in much more than a third of the published articles. This poor replication rate suggests that the application of VBM to DTI data may not be the optimal way for finding the subtle microstructural abnormalities suggested in schizophrenia.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Schizophrenia is a disease characterized, in part, by abnormalities in white matter that affect brain connectivity (Davis et al., 2003). It is not clear, however, whether these abnormalities affect the entire white matter equally, or if they are localized and limited to specific structures or connections. Both of these hypotheses have been suggested, and are frequently tested using a wide variety of in vitro and in vivo methods. Among those methods is diffusion tensor imaging (DTI), the first and so far the most powerful in vivo method that can be used to quantify the integrity of white matter (by means of the properties of water diffusion in human brain tissue) (Basser et al., 1994). DTI measures the diffusion of water molecules throughout the brain, including water in the cerebrospinal, intracellular, and extracellular fluids. Since ventricular spaces, gray matter, and white matter each have different structural properties, water diffusion in each respective region is likewise unique.

One measure that is of particular interest in DTI studies is fractional anisotropy (FA). FA is a measure of directional preference of water diffusion (Basser, 1995). As the axon bundles and myelin sheaths in the major fiber bundles of the brain cause water to preferentially diffuse parallel with the direction of the fiber bundle, FA can be interpreted as a measure of the integrity of the white matter in the area being investigated. Many different post-processing methods have been utilized to look at changes in FA in schizophrenic patients, with hopes of linking white matter abnormalities to the clinical symptoms seen in this devastating disease.

Post-processing and analytic techniques used for DTI data usually employ one of two approaches. The first approach is hypothesis-driven, where investigators focus on one or a few specific regions of the brain, defined by either manually placed regions of interest (ROIs), or extracted through a more automated means such as fiber tractography. The second approach is hypothesis-free and involves searching the entire brain in order to find regions (voxels) that differ between groups.

Articles taking the first approach have reported both global (or widespread) (e.g. Lim et al., 1999; Minami et al., 2003; Kumra et al., 2004; Mittelman et al., 2006) as well as local FA differences. The latter were reported to be limited to frontal connection abnormalities (Kitamura et al., 2005), interhemispheric connection abnormalities, such as the splenium of the corpus callosum (Foong et al., 2000), and fronto-temporal connection abnormalities, such as the cingulum bundle (Kubicki et al., 2003; Sun et al., 2003), uncinate fasciculus (Kubicki et al., 2002; Price et al., 2008), inferior occipito-frontal
fasciculus (Rosenberger et al., 2008), superior longitudinal fasciculus (Jones et al., 2006), inferior longitudinal fasciculus (Ashtari et al., 2007), fornix (Zhou et al., 2008, Fitzsimmons et al., 2009), and anterior limb of the internal capsule (Buchsbaum et al., 2006; Oh et al., 2009).

Although the ROI and tractography approaches have produced significant results in terms of group differences, these results lack consistency and reproducibility. As already suggested in many previous reviews (i.e. Kubicki et al., 2007; Van Hecke et al., 2009), these inconsistencies possibly originate from a lack of methodological standards for data acquisition, post-processing, and analysis.

As opposed to hypothesis-based DTI analysis, voxel-based morphometry (VBM) is hypothesis-free and allows for an easy, highly automated, and relatively fast analysis of entire populations of subjects. For these reasons it is currently the most popular post-processing method used to compare DTI data across subjects. In this method, all subject scans are co-registered to a common atlas space. Then, images of patients are compared to images of controls voxel by voxel, and regions with statistically significant differences are reported. A number of DTI studies in schizophrenia have utilized VBM to look at differences in FA between patients and normal controls (see Table 1). Using this method, FA differences can be measured over the entire brain at once and, using a statistical threshold, differences in white matter integrity can be reported.

Despite the advantages of using VBM to look at DTI data, it is unclear how reproducible this method is since, until recently, there had been no systematic analysis of VBM studies in schizophrenia. Ellison-Wright and Bullmore (2009) recently published a meta-analysis of 15 different studies published to date, reporting FA reductions in the left frontal deep white matter and the left temporal deep white matter. For reasons presented below, we chose to use a different method of comparison between DTI VBM publications.

DTI VBM studies, unlike structural VBM studies, do not necessarily feature similar scanning and processing methods. More specifically, a variety of different scan parameters are used, as well as a variety of magnet strengths, image resolutions, numbers of diffusion directions, numbers of repetitions, ways of tensor estimation, and even slight differences in analytic parameters, all of which make it impossible to account for in a statistical meta-analysis comparison, such as the one by Ellison-Wright et al. previously mentioned. For these reasons, we sought to gain a clearer picture of the white matter abnormalities observed in schizophrenia as well as the reproducibility of DTI VBM studies by plotting the coordinates from 23 different articles, each utilizing the VBM analytic technique applied to schizophrenic populations, onto a single atlas and analyzing the distribution of observed abnormalities.

2. Methods

We used the Medline and Ovid engines, searching for articles using “DTI” or “diffusion,” “whole brain” or “voxel based” or “VBM,” and “schizophrenia” as keywords. From among the results of these searches, we chose articles that used SPM for registration and reported reduced FA in schizophrenic patients compared to normal controls. Our review covers 23 articles (13 chronic; 10 first-episode/early-onset), all of which meet these criteria, published between February of 1998 and April of 2010.

For most articles, statistically significant coordinates representing reduced FA were extracted from tables or from the text. For articles with no points listed, coordinates were approximated by comparing the images in the articles showing highlighted regions of statistically significant FA reduction to the Talairach atlas using the Talairach Applet (www.talairach.org) and Sleuth v.1.1 (www.brainmap.org). Coordinates for all articles were initially extracted in their reported atlas space (or Talairach space if no points were listed), but were ultimately converted to Montreal Neurological Institute (MNI) atlas space. For this reason, as well as for reasons of consistency, only studies that used SPM for registration were selected and plotted. This led to the exclusion of articles from our original search query that used a newer voxel-based approach, introduced by Smith et al. (2006), tract-based spatial statistics (TBSS). Talairach coordinates were converted to MNI space using an affine transformation. Using 3D Slicer (www.slicer.org), we plotted the coordinates from the articles onto a single MNI atlas.

Cubes representing each coordinate were generated and labeled depending on whether the study was conducted with chronic or early-onset/first-episode patients. We chose to make each cube model 3 × 3 × 3 voxels in size for the following reasons: 1) MNI atlas voxels are 1 mm³ whereas most DTI data are about 2 mm × 2 mm × 2 mm at best. 2) Smoothing increases the size of clusters. 3) Actual cluster sizes were not reported in many of the articles. Finally, in order to quantify the percentage of studies reporting each white matter area, we used the MNI space white matter atlas (Laboratory of Neuro Imaging, 2008) to count the number of coordinates (models) that overlapped with each of the white matter structures (tracts).

3. Results

The coordinates representing reduced FA in patients with schizophrenia were scattered across the brain (Fig. 1). The genu of the corpus callosum, the splenium of the corpus callosum, the right anterior corona radiata, and the posterior thalamic radiation bilaterally (including the optic radiation) were the most consistently reported regions; however, each was reported in only six (26.09%; genu of the corpus callosum and left posterior thalamic radiation), seven (30.43%; right anterior corona radiata and right posterior thalamic radiation), or eight (34.78%; splenium of the corpus callosum) of the 23 articles reviewed. Other instances of reduced FA were replicated at an even lower rate.

When the results were grouped by time since diagnosis, results improved only marginally. One region was reported in five (38.46%) of the 13 chronic schizophrenia studies: the left posterior thalamic radiation (including the optic radiation). Likewise, first-episode and early-onset patient studies reported two regions in four of the 10 studies: the splenium of the corpus callosum and the right anterior corona radiata. In addition, 16 of the 23 articles (10 chronic; six early-
onset) reported coordinates that were outside of any label on the white matter atlas. Our observations show a need for increased consistency and reproducibility among DTI VBM studies in schizophrenia.

4. Discussion

Results of our investigation indicated that when transferred to a common atlas space, findings from publications utilizing DTI and investigating whole-brain white matter integrity in schizophrenia are shown to be scattered across the brain and largely inconsistent. Only one region, the splenium of the corpus callosum, was reported in eight of the 23 articles reviewed, only two regions were reported in seven, and only two regions were reported in six. In addition, all five of the most commonly reported regions were not reported together in any of the articles. Furthermore, although all of the regions considered here have shown decreased FA for patients, increased FA (Hubl et al., 2004; Seok et al., 2007; Shergill et al., 2007) and negative results (Moriya et al., 2010) have also been reported.

In our set of articles, measures other than FA were used with VBM as well but with a much lower frequency (e.g. mean diffusivity, mentioned in four out of the 23 articles). The reproducibility of the mean diffusivity results does not seem to be higher than that of FA, though there are too few articles in our study that included this analysis to reach a definitive conclusion.

Such inconsistencies of schizophrenia DTI VBM findings and their lack of replications certainly require attention. Understanding the meaning of these findings and the possible source of observed discrepancies will hopefully lead us to better analytic protocols. It may also get us closer to understanding the relationship between anatomy, microstructure, and DTI abnormalities.

When closely reviewing the publications in the field of DTI and schizophrenia, it is apparent that their comparison, as well as their combined meta-analyses, might be quite challenging. There are two major reasons for this challenge: one is related to the populations studied and the other to DTI data itself. Each can influence the results and potentially lead to inconsistencies, such as those we observed in the present review. Below, we discuss some of these potential confounds, related both to the subject sample populations and to the DTI data:

One of the reasons for inconsistencies in the results reported in DTI, as well as in other areas of schizophrenia research, is the heterogeneity in symptoms for people diagnosed with schizophrenia. Current schizophrenia diagnostic methods require patients to show at least two of the following symptoms: delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, or negative symptoms (American Psychiatric Association, 2010). Because the disorders in this list affect a number of different cognitive domains, two patients with completely different cognitive/clinical profiles might both carry the diagnosis of schizophrenia. A few studies that segregated subjects by symptoms already seem to confirm this hypothesis. Such studies have found that certain positive symptoms, such as hallucinations, are linked with increases in FA (Hubl et al., 2004; Seok et al., 2007; Shergill et al., 2007), whereas negative symptoms are more likely to be linked with FA decreases (Bai et al., 2009). Some studies also show decreased FA in poor outcome patients when compared to good outcome patients (Mitelman et al., 2007). Therefore white matter changes may be more specific to symptoms, making it difficult to draw conclusions from studies in which populations consist of patients with heterogeneous symptoms. Studying homogeneous populations, on the other hand, would lead to an increased understanding of which patient symptoms cause different observed DTI findings (Shenton et al., 2001).

Apart from different clinical characteristics, populations differ with respect to age, gender, race, and medication, to name just a few. Unfortunately, most, if not all, of these factors are believed to be related to white matter pathology. For example, there are indications that white matter might mature differently in individuals with schizophrenia than in healthy controls (Tang et al., 2009), leading to difficulties in interpreting DTI data when subjects are not matched accordingly (Jones et al., 2006; Rosenberger et al., 2008). Medication may also be a factor in...
differing white matter properties across subjects. Diffusivity has been shown to decrease after medication, but only if the patients were drug-responders; no significant diffusivity differences were seen in poor responders before and after medication (Garver et al., 2008). Positive correlations between antipsychotic dose and FA have also been found in both the left frontal white matter and the right middle cerebellar peduncle (Minami et al., 2003; Okugawa et al., 2004).

All the examples discussed above, along with the results of our current investigation, suggest that the white matter pathologies detected with DTI might be specific to the studied population, rather than to schizophrenia in general. One solution involves significantly increasing the sample size in order to discover white matter changes common to all of the schizophrenia subpopulations. The alternative solution requires identifying clinical subpopulations and analyzing each of them separately.

As previously mentioned, the second major possible factor responsible for the lack of reproducibility among schizophrenia DTI VBM studies may arise from differences in the DTI data itself. These include differences in data acquisition, processing, and analysis. For example, data resolution has been shown to affect FA in white matter (thinner slice resolution was related to higher FA values) (Oouchi et al., 2007). Diffusion weighting (or B value) has been shown to impact FA values (Alexander and Barker, 2005), and different B values have been shown to be sensitive to different water compartments (Ben Bashat et al., 2005). Data processing also has an enormous impact on study results. Even though no parametric studies have been performed to understand all possible implications of processing parameters on data analysis, differences in smoothing (smoothing minimizes, to some degree, registration errors in VBM studies), for example, have been shown to produce very different results depending on the filter size (Jones et al., 2005). Inter-subject registration and the way by which it is performed in voxel-based approaches also seems to affect group differences in clinical results, as suggested by the example of two TBSS analyses of the same data set showing much more extensive changes when more precise registration (a newer software version) was employed (Asami et al., 2009).

It is worth mentioning that when the results from all schizophrenia investigations were plotted together, white matter abnormalities in schizophrenia seemed to involve almost the entire white matter. It is thus possible that because of the aforementioned confounding factors (clinical, anatomical, and methodological), DTI derived measures are not sensitive and specific enough to detect possibly subtle white matter abnormalities in schizophrenia, especially at such small sample sizes. If this is the case, results observed in our study could be explained by the fact that when larger datasets are analyzed together (such as in this case), the confounding factors discussed above play a smaller role and common abnormalities that characterize the entire schizophrenia population are being detected.

Besides the biological explanation of observed whole-brain DTI VBM abnormalities in schizophrenia, there is another explanation that is related to the methodology of the VBM studies reviewed here. This issue is related to the subject-to-subject co-registration of images and problems related to this approach when it is applied to schizophrenic populations, where neurodevelopmental changes might occur. Voxel-based morphometry assumes anatomical conformity among subjects. However, even healthy brains are characterized by moderate anatomical variability involving various gray matter structures and, as an effect, the architecture of white matter connections between these gray matter regions. Furthermore, these anatomical differences might be exaggerated in schizophrenia, which is viewed as a neurodevelopmental disease with possible abnormalities involving cortical complexity (Crow et al., 1989; Matsumoto et al., 2001). These slight anatomical differences in gray and white matter architecture between controls and schizophrenics can cause false positive findings as registration attempts to warp images into conformity. Such errors, though they indicate localized differences in studies involving small samples, when put together, could potentially produce results similar to those observed here. In this scenario, the results of this investigation might be indicative of abnormalities in the brain architecture, rather than in the white matter microstructure.

Our hypotheses need to be confirmed by further studies. If indeed the entire white matter is involved in schizophrenia, analyses of larger datasets and the combination of DTI with other imaging methods (such as MTR, spectroscopy, etc.) should be performed in order to confirm this finding. In addition, more attempts to differentiate between white matter abnormalities affecting the macro-architecture versus the microstructure are needed. Here, more precise anatomical tools for DTI data analysis, including diffusion tractography and tract parametrization (e.g. Maddah et al., 2008), give us a better chance to observe abnormalities along specific anatomical structures and should be used.

We believe that our method of plotting the coordinates on a single atlas was appropriate because it allowed for direct comparison of the results from different studies. However, in analyzing these results, one cannot simply compare the average distance between coordinates with that of a random sprinkling of points on an atlas because that assumes FA is normally distributed across the brain. Fiber crossing and fanning ensure that, even across the white matter, FA is not normally distributed. In any case, reproducibility of future VBM studies may be increased by standardized methodological procedures (Van Hecke et al., 2009).

A comprehensive analysis of the results of diffusion tensor voxel-based studies that have been conducted to date in schizophrenia demonstrates that the results are inconsistent, non-replicated, and scattered over the entire white matter. Such diversity among findings most likely arises from both clinical and methodological variabilities that are usually very difficult to control for. Cumulative results of this investigation, however, do suggest several possible ways for improving such studies. The first conclusion is that the sample sizes that are being used in clinical studies are likely too small. Clinical and anatomical variability among subjects, as well as the lack of methodological standards, might require large sample sizes for DTI to be methodologically sensitive enough to detect subtle brain abnormalities. Secondly, the anatomical variability (differences in the shape, thickness, length, number, and location of both individual axons as well as entire fiber tracts) already observed in healthy controls is possibly exaggerated in schizophrenia. This casts doubt on the validity of the assumption—that it is possible to achieve perfect anatomical matching through subject-to-subject registration (even with ever-improving registration algorithms)—behind VBM and other similar whole-brain DTI analytic methodologies. Also, slight modifications in the analytic pipeline tend to completely change the study outcome, and the non-normal distribution of FA puts into question the sensitivity of voxel-based approaches for studying local microstructural abnormalities. Other analytic methods, or a combination of DTI with other imaging and non-imaging methods, should be used in this case.

Acknowledgments

This work was supported by the National Institute of Health (K05 MH070047 and R01 MH 50740 to MES, R01 MH 082918 to SB and R01MH 074794 to MK, P50MH 080272 to MES), the Department of Veterans Affairs Merit Awards (MES), the VA Schizophrenia Center Grant (MES). This work is also part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149 (MK, MES).

References
