Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2016 February 27; 9788: . doi:10.1117/12.2216517.

A framework for incorporating DTI Atlas Builder registration into Tract-Based Spatial Statistics and a simulated comparison to standard TBSS

Matthew Leming^a, Rachel Steiner^c, and Martin Styner^{a,b}

^aDepartment of Computer Science, University of North Carolina, Chapel Hill, NC 27599 USA

^bDepartment of Psychiatry, University of North Carolina, Chapel Hill, NC 27599 USA

^cVanderbilt University, Nashville, TN 37232

Abstract

Tract-based spatial statistics (TBSS)⁶ is a software pipeline widely employed in comparative analysis of the white matter integrity from diffusion tensor imaging (DTI) datasets. In this study, we seek to evaluate the relationship between different methods of atlas registration for use with TBSS and different measurements of DTI (fractional anisotropy, FA, axial diffusivity, AD, radial diffusivity, RD, and medial diffusivity, MD). To do so, we have developed a novel tool that builds on existing diffusion atlas building software, integrating it into an adapted version of TBSS called DAB-TBSS (DTI Atlas Builder-Tract-Based Spatial Statistics) by using the advanced registration offered in DTI Atlas Builder⁷. To compare the effectiveness of these two versions of TBSS, we also propose a framework for simulating population differences for diffusion tensor imaging data, providing a more substantive means of empirically comparing DTI group analysis programs such as TBSS. In this study, we used 33 diffusion tensor imaging datasets and simulated group-wise changes in this data by increasing, in three different simulations, the principal eigenvalue (directly altering AD), the second and third eigenvalues (RD), and all three eigenvalues (MD) in the genu, the right uncinate fasciculus, and the left IFO. Additionally, we assessed the benefits of comparing the tensors directly using a functional analysis of diffusion tensor tract statistics (FADTTS¹⁰). Our results indicate comparable levels of FA-based detection between DAB-TBSS and TBSS, with standard TBSS registration reporting a higher rate of false positives in other measurements of DTI. Within the simulated changes investigated here, this study suggests that the use of DTI Atlas Builder's registration enhances TBSS group-based studies.

Keywords

White Matter; Tract-Based Spatial Statistics; Registration; Diffusion Tensor Imaging; DTI Atlas Builder

1. DESCRIPTION OF PURPOSE

Tract-Based Spatial Statistics is a program widely used in group population studies of DTI² datasets. Nonetheless, as many other groups have noted¹, TBSS suffers from its lack of use of directional data in all stages of its analysis, opting, in most cases, to examine Fractional

Anisotropy instead. But besides criticisms of registration techniques, most criticisms of TBSS have lacked consistency with one another^{1,5}; in part, this is because there is no generalized way in which to determine whether other techniques are superior, with most studies relying on real datasets without ground truth. This study hopes to address these issues. First, we present DTI Atlas Builder-Tract-Based Spatial Statistics (DAB-TBSS), which uses the registration method employed in DTI Atlas Builder (via ANTS-SyN) in the atlas building steps of TBSS. We then present a simulation study of group-wise DTI properties, applied to both DAB-TBSS and TBSS, in order to determine which is able to better capture these synthetic differences. We use measurements of DTI other than FA (which very few studies⁴ have focused on in the past) in the end in order to assess which is generally the best in group difference detection. As a follow up, we also explore the effectiveness of comparing entire tensors in DAB-TBSS as opposed to scalar measurements of tensors.

2. METHODS (DAB-TBSS)

The DAB-TBSS processing workflow is similar to the one of TBSS (the full guide to TBSS can be found in the TBSS user guide (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS/UserGuide>) as visualized in Figure 1. To make the organization of the results easier, the overall setup for folders and naming conventions in DAB-TBSS have been kept the same. This section will describe the entire proposed DAB-TBSS from beginning to end for clarity, though beyond the atlas generation, this methodology is identical to TBSS.

2.1 Pre-processing

DAB-TBSS takes as input DTI image sets. These sets must first be skull stripped.

2.2 Atlas population registration

We register each case from the population to a common atlas using DTI Atlas Builder⁷. To summarize this process, we generate intensity calibrated FA images, built first with an affine followed by an unbiased, diffeomorphic fluid deformation based atlas, finalized by remapping the data via ANTS-SyN. After each successful registration step, the deformation transforms are applied to the original tensor images and the iteratively improving atlas is computed in tensor space. A final brain mask generated in the atlas space is applied to each of the atlas-mapped images.

2.3 Generate measurements of DTI

With all DTI datasets registered into the same space, FA, AD, MD, and RD are estimated from each of the DTI sets. While measurements other than FA are normally left as options to the user, they are performed automatically in DAB-TBSS.

2.4 White matter skeleton estimate and projection

A white matter skeleton is estimated from the FA atlas (the full procedure for this is described in⁶), in order to act as a representation of the centers of the white matter tracts in a dataset; this step also estimates the vectors perpendicular to each voxel of the skeleton. To achieve fine alignment, the maximum FA value of each subject underneath these vectors is

projected onto the skeleton, rather than the FA value directly underneath the skeleton. If the user specifies, they may choose to project the AD, RD, or MD values onto the skeleton instead, but the vector will still be generated by the highest FA value. This gives voxel-by-voxel values over which each dataset may be compared. From this, a statistical analysis, such as multiple regression, may be applied across subjects to find an average difference between two groups for each voxel. Note that each measurement of DTI has to be processed separately.

2.5 Viewing results

We are visualizing DAB-TBSS results via the standard TBSS viewer in FSL, called FSLVIEW³, see Figure 2.

3. SYNTHETIC SIMULATION STUDY

A problem with TBSS, or any such group comparison software, is quantifying whether any potential alteration to the pipeline is an improvement or not. In most cases^{1,5}, alterations are tested on datasets acquired by the research groups that conducted the study. However, since these datasets are usually typical to a particular kind of study, they are a poor way to test differences in the general case. If both TBSS and its proposed alterations find a false positive or false negative at a point along the dataset, it is necessary to point out a common weakness in both methods.

To address this, we employed a Matlab-based simulation for group effects that creates synthetic differences in a given set of diffusion tensor images. First, a user inputs a binary mask for each dataset, indicating which tensors they would like to alter. Then, the eigenvalues of tensors under the mask are altered via scaling with a normally distributed scaling factor. The first eigenvalue, second eigenvalue, third eigenvalue, or a combination thereof may be altered. Optionally, the user may also specify mean and standard deviation values to edit the magnitude of the second and third eigenvalues. In order to maintain the isotropy of factors such as cerebral-spinal fluid, only tensors that exceed a certain fractional anisotropy are altered (we found a threshold FA of 0.1 to be acceptable). This offers ground truth of statistically significant differences between two datasets (the simulated and the original), which the group comparison software then tests.

For this study, we selected DTI datasets of a group of 37 subjects from an intrauterine exposure study on drug-abusing mothers, acquired on a 3T Siemens scanner with 21 unique diffusion directions at $b=1000$. Four of these 37 were taken out of the study during quality control of the raw diffusion data, making for a total of 33. These 33 were simulated to have differences in selected regions in right uncinate fasciculus, genu, and left inferior fronto-occipital fasciculus (IFO). To generate masks for these fibers, we estimated the fiber tracts of the genu, the right uncinate, and the left IFO for all subjects, generating a mask covering all voxels through which one or more tracts passed. Using these masks for each subject, we performed simulations using the following parameters for Gaussian random values: $(\mu=1.005, \sigma=0.001)$; $(\mu=1.01, \sigma=0.001)$; $(\mu=1.02, \sigma=0.001)$; $(\mu=1.05, \sigma=0.001)$; $(\mu=1.10, \sigma=0.001)$; $(\mu=1.15, \sigma=0.001)$, for the principle eigenvalue (altering AD), the second and third eigenvalues (altering RD), and all three eigenvalues (altering MD), making a total of 30

different simulations. Results from this simulation will allow us to look at varying magnitudes of the group difference for the principal eigenvalue. Thus, 33 real subjects were compared to their 33 simulated counterparts, and no other differences outside those simulated differences should be predicted to occur between these groups.

The number of voxels in the final atlas that achieved a p-value under 0.05 were counted using three hand-drawn masks over each of the respective areas in which the altered fibers were located. Because the genu, right uncinate, and left IFO are sufficiently far away from each other and leave little risk of intersection (i.e., detecting a change in fiber A that is actually present in fiber B), this method was considered appropriate.

4. FULL TENSOR COMPARISON

In addition to comparing scalar measurements of diffusion tensors, we also compared the tensors directly (Figure 4), using procedures from an existing DTI statistical analysis software performing functional data analysis of diffusion tensors^{8,10}. Tensors from voxels that were projected onto the skeleton in the final step of TBSS were instead compared across all groups for a final p-value. Due to the heavy computational aspect of this task, we opted for 1,000 simulations instead of the 10,000 used to compare scalar measurements. This procedure was performed for DAB-TBSS results as a way to test ways to increase sensitivity, but it was not tested for regular TBSS due to software limitations. We include these results in this study not as a way to directly compare DAB-TBSS and TBSS, but as a suggestion of future directions for TBSS.

5. RESULTS

We assessed the effectiveness of DAB-TBSS and TBSS, first, by its ability to detect differences in scalar measurements between datasets. We then assessed the presence of false positives detected across all simulations. In general, DAB-TBSS picked far fewer false positives than TBSS in areas where TBSS was more sensitive to detecting change. DAB-TBSS using direct tensor comparisons gave the most sensitive and accurate results. However, excluding the direct tensor comparisons (which is an unfair comparison between DAB-TBSS and TBSS, since direct tensor comparisons were not tested on standard TBSS), FA and AD measurements appeared to give the most favorable results in these tests.

The following charts show the results of the three simulations (altering AD, MD, and RD) with the lowest simulated percent group difference at which at least one voxel in the fiber is detected at $p=0.05$, i.e. at least one true positive location.

5.1 True positives

5.2 False positives

This next set of tables shows the smallest alterations for which false positives were detected

- a When any significant location appeared in DTI properties other than the simulated/affected property and thus would be expected to be unaffected by the simulated differences. In other words, if AD were increased in a test group, one

would expect the FA, AD, and MD for the test group to be greater than the control group, but RD to be unaffected.

- a** When the detected differences in an expectedly affected DTI property is in the opposite direction of the simulated differences.

In summary, it was found that DAB-TBSS's registration method picked up subtler differences than TBSS for most settings, though TBSS appeared to work more effectively for picking up differences in FA. In addition, DAB-TBSS provided fewer false positives studied here, whereas TBSS provided numerous false positives across the simulations and measurements studied. Regarding scalar measurements of tensors, AD gave the most favorable results, presumably because the tensors were changed based on proportion and AD would thus change the form of the tensor in the most substantial way.

In DAB-TBSS, the tensor comparisons detected differences across all simulations, failing only at the RD simulation in the left IFO and the right uncinate, which were the weakest tracts in the most subtle simulations; the results imply that changes greater than 15% would have led to detection in the RD simulation in this case.

6. DISCUSSION

While it is intuitive that any tensor-based registration method would produce superior results to one that uses only scalar values, it is necessary to develop methods of empirically comparing registration methods in TBSS and similar analysis software. This study is a proof-of-concept of one such framework.

A potential area of improvement in this comparison is an assessment of the areas of the p-values. We compared the amount of change necessary to the eigenvalues of a diffusion tensor to bring the p-value in one voxel to below 0.05; because we are working with ground truth data and are trying to assess the absolute minimal amount of change necessary for either framework to detect change, this was thought to be a good means of presenting the effectiveness of both toolboxes. This also served as a metric that was simple to compute and understand, being invariant to the resolution of the testing data (the output of DAB-TBSS had fewer voxels in the data than the default atlases of TBSS, making a total count of the number of voxels with $p > 0.05$ useless). While more variations between datasets would be seen in real world data, this simple metric does serve to compare techniques when ground truth is known.

Consistent with theory, in DAB-TBSS, the tensor comparisons picked up differences across all simulations. Whereas many scalar measurements across simulations detected simulations with greater sensitivity than the tensors comparisons, the tensor comparisons consistently picked up differences across simulations, whereas particular scalar measurements were not. For example, FA was typically a good indicator of differences; however, in the MD simulations, consistent with theory, FA did not detect any differences, whereas tensor comparisons did. This would suggest that, in a general case in which the nature of fiber differences is unknown, it may be safest to compare tensors to find potentially affected areas before taking scalar measurements to learn more about the nature of these changes.

The future directions of this work would be to assess the effectiveness of different tensor-based registration techniques in TBSS (such as DTI-TK⁹), as well as different means of comparing multiple tensors other than that presented in FADTTS. This will help TBSS become a tensor-based analysis program rather than scalar, and it will allow a way to determine which new software and methodologies can be helped to improve TBSS in this direction.

7. CONCLUSION

In this comparison, we have used tests on a ground truth to argue that the default registration technique used in TBSS is comparable in its detection rates to the ANTS-SyN-based registration employed in DTI Atlas Builder, but is more prone to error, especially for non-FA measurements of diffusion tensors. We further argue that it is necessary to use a ground truth comparison framework in order to assess the effectiveness of different methodologies.

It was found that the proposed alterations of TBSS were more effective at detecting simulated differences. It was found that testing for these simulated differences directly on AD values in DAB-TBSS successfully detected these changes at a 5–10% magnitude of change, without producing a false positive throughout the dataset. While standard TBSS detected these differences, it also produced a number of false positives throughout the dataset. The FA measurement on TBSS found differences in the 10–15% range without such false positives. Thus, the method, in this study, that found the simulated differences most accurately, was DAB-TBSS with AD measurements. This suggests that non-FA measurements in TBSS are useful, but the standard TBSS registration method may not lead to group difference detection as accurate as the one in DTI Atlas Builder.

References

1. Bach M, Laun FB, Leemans A, Tax CMW, Biessels GJ, Stieltjes B, Maier-Hein KH. Methodological Considerations on tract-based spatial statistics (TBSS). *Neuroim.* 2014; 100:358–369.
2. Bassler PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophysical journal.* 1994; 66(1):259–267. [PubMed: 8130344]
3. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage.* 2012; 62(2):782–90. [PubMed: 21979382]
4. Nave RD, Ginestroni G, Diciotti S, Salvatore E, Soricelli A, Mascalchi M. Axial diffusivity is increased in the degenerating superior cerebellar peduncles of Friedreich's ataxia. *Neurorad.* 2011; 53:367–372.
5. Schwarz CG, Reid RI, Gunter JL, Senjem ML, Przybelski SA, Zuk SA, Whitwell JL, Vemuri P, Josephs KA, Kantarci K, Thompson PM, Petersen RC, Jack CR. Improved DTI registration allows voxel-based analysis that outperforms Tract-Based Spatial Statistics. *Neuroim.* 2014; 94:65–78.
6. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroim.* 2006; 31(4):1487–1505.
7. Verde AR, Budin F, Berger J, Gupta A, Farzinfar M, Kaiser A, Ahn M, Johnson H, Matsui J, Hazlett HC, Sharma A, Goodlett C, Shi Y, Gouttard S, Vachet C, Piven J, Zhu H, Gerig G, Styner M. UNC-Utah NA-MIC framework for DTI fiber tract analysis. *Frontiers in Neuroinformatics.* 2014; 7(51)
8. Yuan Y, Zhu H, Styner M, Gilmore JH, Marron JS. Varying coefficient model for modeling diffusion tensors along white matter tracts. *Ann Appl Stat.* 2013; 7(1):102–125. [PubMed: 24533040]

9. Zhang H, Yushkevich PA, Alexander DC, Gee JC. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis*. 2006; 10(5):764–785. [PubMed: 16899392]
10. Zhu H, Kong L, Li R, Styner M, Gerig G, Lin W, Gilmore JH. FADTTS: Functional analysis of diffusion tensor tract statistics. Elsevier. 2011; 56(3):1412–1425.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

The overall processing scheme of DAB-TBSS. Skull-stripped images are input into DTI Atlas Builder, then the FA of these is used to estimate the white matter skeleton. The maximal FA value in a perpendicular range of each registered image is projected onto the FA skeleton. RD, AD, and MD are automatically estimated from the DTI and projected in their own datasets using the FA vector onto the white matter skeleton. For each voxel, a group difference algorithm is employed. FSLVIEW displays the results.

Figures 2 and 3.

TBSS (2) and DAB-TBSS (3) results in the AD simulations at ($\mu=1.005$, $\sigma=0.001$); ($\mu=1.01$, $\sigma=0.001$); ($\mu=1.02$, $\sigma=0.001$); ($\mu=1.05$, $\sigma=0.001$); ($\mu=1.10$, $\sigma=0.001$); ($\mu=1.15$, $\sigma=0.001$), in the left IFO. Shown is the differences detected by measuring FA, displayed here at voxels for which $p < 0.10$. Statistically significant differences shown by the red-yellow voxels.

Differences in slices are due to TBSS's use, in the registration step, of a higher-dimensional, predefined atlas for registration, while DTI Atlas Builder takes an average of all datasets, preserving the original dimensions of an image. Slices are adjusted to show where differences occurred in the left IFO.

Figure 4.

A frontal view of the genu in simulations in which the principal eigenvalue was increased by 15%. From left to right, the detections of (1) the tensor comparison, (2) the AD measurement, (3) the FA measurement, and (4) the MD measurement. Red to yellow displays the range of $0 < p < 0.05$, with yellow signaling lower p-values. Slices are displayed to show those areas around the genu with approximately the most detected differences detected; thus, slices in the view may not be the same, but they do display the detections between different forms of measurements the most accurately.

Table 1

Tensor comparison: Minimum percent change in each of the simulations (to AD, RD, and MD) at which the tensor comparison detected a statistically significant difference between groups ($p < 0.05$).

	Genu	Right Uncinate	Left IFO
AD	5%	15%	15%
MD	10%	10%	15%
RD	15%	–	–

Table 2

Alterations to AD: Minimum percent change to the principal eigenvalue for which each program detected at least one pixel when looking at the following scalar measurements:

		Genu	Right Uncinate	Left IFO
DAB-TBSS	FA	10%	10%	15%
	AD	5%	10%	10%
	MD	10%	15%	10%
	RD	–	–	–
TBSS	AD	1%	–	–
	FA	10%	10%	10%
	MD	0.5%	–	–
	RD	–	–	–

Table 3

Alterations to MD: Minimum percent change to the principal, second, and third eigenvalues for which each program detected at least one pixel when looking at the following scalar measurements:

		Genu	Right Uncinate	Left IFO
DAB-TBSS	FA	–	–	–
	AD	5%	10%	10%
	MD	10%	10%	10%
	RD	15%	–	–
TBSS	FA	–	–	–
	AD	–	–	–
	MD	5%	–	–
	RD	15%	–	–

Table 4

Alterations to RD: Minimum percent change to the second and third eigenvalues for which each program detected at least one pixel when looking at the following scalar measurements:

		Genu	Right Uncinate	Left IFO
DAB-TBSS	FA	10%	15%	15%
	AD	–	–	–
	MD	15%	–	–
	RD	–	–	–
TBSS	FA	10%	10%	10%
	AD	0.5%	–	–
	MD	0.5%	–	–
	RD	1%	–	–

Alterations to AD: The errors detected within the indicated fiber regions at each percent change in the simulations to the principal eigenvalue, as well as any statistically significant differences detected outside of any of the areas of change.

Table 5

	Genu	Right Uncinate	Left IFO	Outside fibers
DAB-TBSS	FA- type b	–	–	–
	AD- type b	–	–	10–15%
	MD- type b	–	–	–
	RD- type a	–	–	–
TBSS	FA- type b	–	–	–
	AD - type b	0.5–15%	2%, 10–15%	0.5–15%
	MD- type b	15%	10–15%	0.5–15%
	RD - type a	0.5–15%	10–15%	0.5–15%

Table 6

Alterations to MD: The errors detected within the indicated fiber regions at each percent change in the simulations to the principal, second, and third eigenvalues, as well as any statistically significant differences detected outside of any of the areas of change.

		Genu	Right Uncinate	Left IFO	Outside fibers
DAB-TBSS	FA- type a	–	–	–	–
	AD- type b	–	–	–	10–15%
	MD - type b	–	–	–	–
	RD- type b	–	–	–	–
TBSS	FA- type a	–	–	–	–
	AD- type b	2–15%	–	–	5–15%
	MD - type b	–	–	–	2–15%
	RD- type b	2–15%	–	–	2–15%

Alterations to RD: The errors detected within the indicated fiber regions at each percent change in the simulations to the second and third eigenvalues, as well as any statistically significant differences detected outside of any of the areas of change.

Table 7

		Genu	Right Uncinate	Left IFO	Outside fibers
DAB-TBSS	FA- type b	–	–	–	–
	AD- type a	–	–	–	–
	MD- type b	–	–	–	–
	RD- type b	–	–	–	–
TBSS	FA- type b	1%,5–15%	–	–	15%
	AD- type a	0.5–15%	15%	–	0.5–15%
	MD- type b	–	10%–15%	–	0.5–15%
	RD- type b	0.5–15%	15%	–	0.5–15%