Title:
Methodological considerations on Tract-Based Spatial Statistics (TBSS)

Author names and affiliations:

Michael Bach[1,2], Frederik B. Laun[1,2], Alexander Leemans[3], Chantal M.W. Tax[3], Geert Jan Biessels[4], Bram Stieltjes[1], Klaus H. Maier-Hein[1,5]

1: Section Quantitative Imaging-based Disease Characterization, Department of Radiology, German Cancer Research Center (DKFZ)

2: Department of Medical Physics in Radiology, German Cancer Research Center (DKFZ)

3: Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

4: Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, the Netherlands

5: Computational Disease Analysis Group, Div. Medical and Biological Informatics, German Cancer Research Center (DKFZ)

Corresponding author:

Dr. rer. nat. Klaus H. Maier-Hein (né Fritzsche)

Computational Disease Analysis Group

Div. Medical and Biological Informatics (E130)

German Cancer Research Center (DKFZ)

Im Neuenheimer Feld 280

69120 Heidelberg, Germany

Email: k.maier-hein@dkfz-heidelberg.de

Phone: (+49) 6221/42-3545

Fax: (+49) 6221/42-2345

Abstract

Having gained a tremendous amount of popularity since its introduction in 2006, Tract-Based Spatial Statistics (TBSS) can now be considered as the standard approach for voxel based analysis (VBA) of diffusion tensor imaging (DTI) data. Aiming to improve the sensitivity, objectivity, and interpretability of multi-subject DTI studies, TBSS includes a skeletonization step that alleviates residual image misalignment and obviates the need of data smoothing. Although TBSS represents an elegant and user-friendly framework that tackles numerous concerns existing in conventional VBA methods, it has limitations of its own, some of which have already been detailed in recent literature. In this work, we present general methodological considerations on TBSS and report on pitfalls that have not been described previously. In particular, we have identified specific assumptions of TBSS that may not be satisfied under typical conditions. Moreover, we demonstrate that the existence of such violations can severely affect the reliability of TBSS results. With TBSS being used increasingly, it is of paramount importance to acquaint TBSS users with these concerns, such that a well-informed decision can be made whether and how to pursue a TBSS analysis. Finally, in addition to raising awareness by providing our new insights, we provide constructive suggestions that could improve the validity and increase the impact of TBSS drastically.

Keywords: TBSS, pitfalls, DTI, quantitative, FA, evaluation

1. Introduction

Diffusion magnetic resonance imaging (MRI) can deliver insight into the living human brain in health and disease, especially in white matter anatomy, and provides quantitative parameters related to white matter (WM) microstructure (Tournier et al., 2011). Much of the knowledge on changes in WM microstructure that we have gained from diffusion MRI originates from studies that compared such diffusion markers between populations of interest, commonly a healthy control group and a diseased population. The value and impact of such studies is directly tied to the ability of researchers to present results that are unbiased, objective, and anatomically specific. Tract-based spatial statistics (TBSS) (Smith et al., 2006) has become a very popular tool for the evaluation of diffusion tensor imaging (DTI) data in this context.

TBSS pioneered the idea of projecting volumetric data onto a WM skeleton to circumvent the partial volume effect (PVE) and gain statistical power from this dimensionality reduction (Smith et al., 2006). The approach does not require data smoothing and could alleviate many concerns that were raised regarding the conventional voxel based morphometry (VBM) framework that was previously used in many DTI studies (e.g., Jones et al. (2005)). Although TBSS has advanced the state of the art in diffusion MRI group studies significantly, the increased complexity by adding the skeletonization step reduces the overall transparency. In other words, while TBSS is very user-friendly, and delivers comprehensive images, it may also obscure several aspects of the raw data that the reader of a study or even the researcher that performed the analysis might not be aware of. With more and more scientists adopting to the technique, it is therefore increasingly important to raise awareness of the limitations of the approach. In previous studies, some problems related to TBSS have been investigated. Edden and Jones (2011) reported that the shape of the skeleton as well as the statistical results are rotationally variant. Zalesky (2011) quantitatively assessed the performance of the projection algorithm in moderating registration misalignments and showed that only 10% of post-registration misalignment was corrected by the TBSS projection algorithm. Keihaninejad et al. (2012) demonstrated the dependence of specificity and sensitivity of TBSS results on the registration target and suggest the use of a group-wise atlas as target. Van Hecke et al. (2010) discussed potential pitfalls and limitations of TBSS, like the assumption that the effect of interest occurs in voxels where the local FA is highest. In the following, we discuss important issues that we address in this study.

One major point of debate is the potentially limited anatomical specificity of TBSS. The technique was introduced as being “tract-based”, in response to the challenge of comparing voxels of “the same part of the same WM tract from each and every subject”, both “in terms of resolving topological variabilities and in terms of the exact alignment of the very fine structures present in such data” (Smith et al., 2006). However, the distinction between adjacent, differently oriented fiber bundles with similar FA values is challenging and alternative methods are described by Kindlmann et

al. (2007) and Yushkevich et al. (2008) to overcome this limitation. Since TBSS only makes use of the FA map and discards the orientation information captured in the diffusion data, two different problems arise. First, complications in terms of anatomical specificity occur in regions where pathways of different structures merge, such as those related to the superior projections of the corpus callosum (CC) and the corona radiata fiber bundles. Without the (long-distance) directional tract information derived from the orientation information, it is virtually impossible to assign the FA values to the same anatomical structure across subjects in a consistent way as the skeletonization step causes these different bundles to collapse on top of each other (see Fig. 1). Furthermore, even in regions where the assignment of voxels to tracts is unambiguous, the tract-specificity of the TBSS projection step is unknown. The region where the cingulum bundle (CB) and CC are in close proximity is a good example in this respect and in the original TBSS-paper, it has been explicitly stated that the CB and CC are correctly differentiated by the projection algorithm (Smith et al. (2006), page 1494, second paragraph): “The superior part of the cingulum (i.e., above the corpus callosum) is slightly extended across its cross-section in the inferior-superior direction, and well-localised across subjects by virtue of the strong, nearby corpus callosum, and hence the normal projections described above work well (similar issues relate to the fornix)”. However, this was not shown experimentally. Since we question the tract-specificity of TBSS throughout this paper, we do not use the words “tract-center” or “tract” when referring to the skeleton, but “locally maximal FA value”, or “FA-skeleton”, because we think this is a less ambiguous and, thus, more appropriate expression.

Another factor that plays a central role in the TBSS processing pipeline, and one that may greatly affect the anatomical specificity of TBSS, is the quality of image registration. The mean FA skeleton has been shown to be less “alignment-invariant” than anticipated and alternative skeleton-based approaches that try to address this point have been published, but have not yet reached a comparable level of acceptance (Kindlmann et al., 2007; Yushkevich et al., 2008; Zhang et al., 2010a).

A further point of debate is the robustness and interpretability of TBSS results. The original TBSS paper includes inter-subject and inter-session test-restest results regarding the reproducibility of FA values (Smith et al., 2006). However, the influence of the user in terms of parameter settings and the noise level on the final TBSS result, i.e., the significant maps, has not been shown. Being a fully automated technique, TBSS is generally considered to be largely user-independent. However, there are several parameters that have to be adopted in each TBSS analysis. While this is potentially very important in order to allow for a proper adaptation of the method to each specific analysis, many papers vary the parameters without motivating their choice. This is critical, since important aspects of the underlying data such as SNR or alignment problems remain unnoticed when looking only at the final result. We anticipate that the influence of different TBSS configuration options on the final result is largely unclear and/or underestimated by TBSS users. One important example is the choice

of template in TBSS studies. Many studies use the FMRIB-template that is distributed with TBSS. This might be mainly due to computational reasons, since the generation of a study-specific target is computationally expensive to obtain, especially in larger populations. However, while the choice of the template is known to significantly impact the results of other group analysis methods (Van Hecke et al., 2011), its impact on the final TBSS result is largely unknown. An initial study was performed by Keihaninejad et al. (2012), who demonstrated the positive impact of improved alignment on TBSS by the use of a group-wise atlas construction.

Taken together, although TBSS may provide plausible results, the final significance maps overlaid on the template image may also hide potential methodological imperfections related to the quality and/or the analysis of the data. In this paper, a deeper look underneath the surface of the TBSS framework is provided. We address several methodological aspects of the technique: how unbiased, objective, and anatomically specific are TBSS results? What are major sources of bias, user-dependence, and non-specificity and to what extent do these factors affect the final TBSS result? With the detailed analyses presented in this study, we provide an in-depth investigation of the major pitfalls when analyzing and interpreting data with TBSS. We conclude with suggestions that define good practice when using TBSS and we propose improvements that may further raise the validity and impact of TBSS.

2. Methods:

2.1 TBSS settings

In all experiments, the TBSS pipeline was applied using the recommended parameters. For the in vivo datasets a permutation test with n=5000, corrected for multiple comparisons and threshold free cluster enhancement (TFCE (Smith and Nichols, 2009)) was used to compare patients and controls, with p=0.05 as threshold for significance. Unless otherwise stated, an FA threshold of 0.2 was applied and the FMRIB58 template (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FMRIB58_FA) was used as registration target. Four dataset types (two in vivo human brain, physical phantom, and synthetic FA images) were used to perform the TBSS analyses in this work. Details on these data sets and experiments are provided in the following paragraphs.

2.2 Dataset types

Two in vivo dataset types were used. The first (*in vivo dataset I*) was acquired at 1.5 T (Symphony, Siemens Medical Solution, Erlangen, Germany) for 15 Alzheimer's disease patients and 15 healthy controls using a twice refocused single-shot echo planar imaging (EPI) sequence. Parameters: repetition time (TR) 4700 ms, echo time (TE) 78 ms, field of view (FOV) 240 mm, in-plane resolution of 2.5 mm, 50 axial slices of 2.5 mm thickness, 6 gradient directions (b=1000 s/mm^2) and a b=0

s/mm^2 image, and 10 repetitions. The second in vivo dataset type (*in vivo dataset II*) was acquired at 3 T (Intera , Philips, Best, the Netherlands) from 50 Alzheimer's disease patients and 50 healthy controls recruited as described previously (Reijmer et al., 2013). A single-shot spin echo EPI with the following parameters was used: TR 6638 ms, TE 73 ms, FOV 220 mm, in-plane resolution of 1.72 mm, 48 axial slices of 2.5 mm thickness, 45 gradient directions (b=1200 s/mm²) and a b=0 s/mm^2 image (number of signal averages=3).

Preprocessing included correction for motion and eddy currents (FSL (Jenkinson et al., 2012), FLIRT (Greve and Fischl, 2009; Jenkinson et al., 2002; Jenkinson and Smith, 2001)) and image masking (FSL, BET (Smith, 2002)). The tensors were estimated with the weighted linear least squares approach (Veraart et al., 2013).

Physical phantom datasets were acquired at 3 T (TRIO, Siemens Medical Solution, Erlangen, Germany) using an EPI sequence with monopolar gradient scheme and the following parameters: Resolution 2.5x2.5x2.5 mm^3, FOV 160x160 mm², TR/TE 2900/78 ms, 180 diffusion directions with b=1000 s/mm^2, 20 b=0 s/mm^2 images, bandwidth=2004 Hz/Px, and GRAPPA acceleration factor 2. All images were corrected for eddy currents using FSL. All voxels with intensities below approximately three times the noise threshold (derived from the mean intensity of the background signal) in the non-diffusion weighted image were excluded from diffusion tensor calculations and are excluded from further analysis.

A *synthetic FA map* of two adjacent WM tracts was emulated by a grayscale image volume with an isotropic voxel size of 1 mm. A small linear FA gradient from each pathway center to its edges is introduced, so that TBSS can identify the center of each trajectory. For a detailed description, see section 2.3.

2.3 Experiments

Effect of WM tract adjacency on anatomical specificity

To investigate tract-specificity, we segmented two major adjacent WM tracts, the CB and the CC, in all subjects in native space (*in vivo dataset I/II*, 15/50 Alzheimer's disease patients and 15/50 healthy controls). We did this by thresholding the main diffusion direction, which is clearly distinguishable between the two tracts, as follows: voxels with FA>0.4 in the analyzed region of interest (Fig. 2a) were marked CB if the first eigenvector deviated not more than 30° from anterior-posterior direction and were defined as part of the CC if they did not deviate more than 30° from left-right direction. Using the same transformations as in the conventional TBSS pipeline, each of the binary segmentations is followed through the TBSS pipeline (Fig. 2b-d).

To determine the potential sources of voxel misassignments between adjacent WM tracts we investigated the effect of the FA skeleton projection procedure and registration quality on the

outcome results. In a first experiment, a conceptual weakness of the TBSS skeletonization and projection step is demonstrated by making use of the *synthetic FA map*. This volume emulates two tracts of different FA (0.9 and 0.6) and different thickness (15 mm and 5 mm) that traverse each other at a 90° angle, similar to the CB and CC (Fig. 3a). The tracts are separated by a 1 mm thick gap. In a second experiment, based on the *in vivo datasets*, we used an alternative registration method (DTI-TK (Zhang et al., 2006)) and repeatedly followed the CB and CC segmentations through the pipeline, in order to assess the influence of registration quality on the misassignment problem. In contrast to the standard TBSS registration (FNIRT), DTI-TK uses the full tensor information for the registration. We choose DTI-TK since it was the overall winner of a registration algorithm challenge as previously published (Wang et al., 2011). Zhang et al. (2007) and Van Hecke et al. (2007) also found that the use of full tensor features or integration of all the diffusion-weighted images instead of tensor-derived indices for the registration can improve the alignment of WM tracts and the detection of WM differences. Because a tensor template is needed as registration target for this approach, the IXI aging template (Zhang et al., 2010b) (65-83 years old, 21 males and 30 females, www.nitrc.org/projects/dtitk) is used for the DTI-TK registration and its FA map for the standard TBSS registration. In a last experiment, in order to evaluate the impact of voxel misassignments on the final TBSS statistical results, we compared the statistical significant maps produced by TBSS at different levels of voxel misassignment (Fig. 4).

Influence of resolution/partial volume and skeleton shape

The skeletonization step of TBSS could – in principle – correct for residual misalignments after the image registration. However, for a successful correction, the direction of the misalignment and the direction of the FA maximum search direction have to match. We analyzed, if this prerequisite for a reliable group comparison is fulfilled in vivo (Fig.5, *in vivo dataset I/II*).

A deeper understanding of the connection between partial volume effects, the skeleton shape, the FA maximum search direction and the projected FA values chosen for the subsequent group comparison is provided by a previously presented resolution phantom (Bach et al., 2013) (*physical phantom datasets*). This phantom (see Fig. 6a) consists of 6 circular fiber strands, each with an outer diameter of 60 mm. They have square cross-sections of 5x5, 3x3, 2.5x2.5, 2x2, 1.5x1.5 and 1x1 mm². Two different image volumes of the phantom were generated by varying its relative position to the imaging matrix by shifting the FOV (Fig. 6b). The two different FOV positions are visualized by the green and red squares in Fig. 6c. In one case, the strand is “halved” by the voxels (green squares). In the other case, the whole strand thickness is covered by just one voxel (red squares). Therefore, the same strand appears in the image with different partial volume effects. This can be seen on the FA maps in Fig. 6b, which shows the six fiber strands of the resolution phantom from the side. In a first

evaluation, we compared the resulting TBSS skeletons that were generated from the different images in order to see whether the TBSS skeletonization step produces consistent results. In a second experiment, we compared the FA values on the skeleton in order to test potential biases that occur during the skeleton projection step of TBSS.

Influence of image noise

Different subsets of the repetitions of the *in vivo dataset I* were used to study the effect of noise level on the TBSS result. In a first evaluation, we varied the number of used repetitions between one (strong influence of noise) to ten repetitions (lowest influence of noise). In a second evaluation, we chose different subsets of two repetitions in order to assess the test-retest capabilities of TBSS. In both evaluations, a standard TBSS-analysis was performed for each of the subsets.

Effect of the user-specified settings

Two parameters that commonly differ between TBSS studies were varied in order to see how strong TBSS results depend on the user input: (i) the choice of the registration target and (ii) the FA threshold defined in the skeletonization process. In a first evaluation, the choice of the registration target was varied between two options, including the FMRIB58 template, which is provided with the TBSS software, and a study-specific target, i.e. the most representative subject from a group of subjects. In a second evaluation, the FA threshold in the skeletonization process was varied between 0.15 and 0.3. The analysis is performed for both *in vivo datasets*.

3. Results

3.1 Influence of adjacent WM tracts

The results in this section were obtained from the *in vivo dataset I/II*. Figures 2e/f show examples of misassigned voxels from the CB to the CC and vice versa (white and black arrows). The contribution of one tract to the other is not binary, even on a voxel-basis, since the registration and interpolation steps introduce a blurring to the binary segmentations. The blue arrows in Fig. 2e/f, for example, point to yellow voxels, where the original colors green and red overlap. Black skeleton voxels represent voxels that could not be identified as being part of the CB or the CC. A quantitative analysis of the subject-specific percentages of voxels that were misassigned is shown in Fig. 2g/h. The analysis differentiates voxels with a contribution, denoted as x, of the wrong WM tract of above 50% and in the range of 10-50%. In summary, 15% of all skeleton voxels had a contribution of 10% or more of the wrong fiber tract. This percentage not only varied strongly across subjects (Fig. 2g), but also between the patient and control groups (Fig. 2h, for $10\% \leq x < 50\%$: $p=0.001$ for dataset I and

p=0.039 for dataset II). In particular, in patients, the overall number of affected skeleton voxels was 25% higher than in controls.

The conceptual limitation of the TBSS skeletonization and projection step is demonstrated using the *synthetic FA map*. Fig. 3a shows the input FA image of two fiber strands and the resulting FA skeleton. The TBSS projection step applies a distance transform to the skeleton (Fig. 3b) in order to determine the search area for the local FA maximum. The resulting search area for the upper skeleton voxel is highlighted in Fig. 3c. It can be seen that the search area partially covers the neighboring fiber tract (Fig. 3d). This is a potential source of misassignment (yellow arrow), especially if the neighboring tract has high FA values. Whenever adjacent WM tract bundles are of different diameter, the search area of the thinner bundle can reach to the thicker bundle and the example demonstrates that even under ideal conditions with a perfect registration and no noise or partial volume, voxels can be misassigned by this procedure. The second source of voxel misassignments that we investigated was the quality of image alignment in the context of tract assignment. To this end, we used the *in vivo datasets*. The overall percentage of voxels, which have a contribution of at least 10% of the neighboring strand, could be reduced by a factor of about seven by replacing the TBSS registration procedure with the DTI-TK image registration approach (Fig. 4b). Fig. 4c demonstrates this effect on the statistical results of TBSS. In particular, conventional TBSS found highly significant group differences in both the CB and the CC. In dataset II this result is more pronounced than in dataset I. However, when using the DTI-TK image registration method, the group differences in the CC became spatially more homogeneous for dataset I and the group differences in the CB disappeared completely for both dataset types. This effect is also reproducible for other registration targets (IXI adult template and study specific template created with DTI-TK; see Supplement 1).

3.2 Influence of partial volume and skeleton shape

Fig. 5a illustrates the mismatch between the FA maximum search direction and the direction of residual registration misalignments in the fornix. Fig. 5b/c show the midsagittal view of the fornix (cyan colored rectangle in Fig. 5a). There is a good concordance between the mean FA skeleton (transparent blue) and the FA maps (background) of the subjects shown in first column of Fig. 5b/c (green arrows). Four further subjects are shown for each dataset and the red arrows indicate areas where the mean FA skeleton does not cover the fornix. The direction for the local maximum search is perpendicular to the skeleton sheet, which lies in the image plane, but the directions of the misalignments between the FA skeleton and the individual FA maps are within this plane (cf. Fig. 5a). In this case, the search for the local FA maxima, which should give the tract center, fails.

Using the *physical phantom datasets*, Fig. 6d/e show the TBSS skeletons for the 2.5 mm thick strand in green and red, respectively. Using TBSS, the fibers were reduced to thin sheets with one voxel thickness. In the first case (green, Fig. 6c and d), this sheet appears thick in the side view and thin in the top view. The second sheet (red, Fig. 6c and e) appears thin in the side view and thick in the top view. The sheet orientation has an impact on the TBSS projection step, since the search direction for the local FA maximum is limited to the directions perpendicular to the sheet. This leads to search directions that are radial to the fiber-ring-plane (Fig. 6d, bottom) in the one case, and perpendicular in the other configuration (Fig. 6e, bottom). Fig. 6f shows a frequency distribution of the FA values derived from the fiber skeleton, demonstrating the influence of the above effect on the projected FA values. In comparison to the red skeleton (red bars), the green skeleton yielded an increased amount of high FA values (green bars). The reason for this effect is the flipping of the search direction of the TBSS projection step that influences the correct identification of FA maxima in the tract center. The FA values on the red skeleton tended to be lower, even though the red configuration was much less effected by partial volume effects than the green configuration.

3.3 Influence of image noise level

The influence of noise level on the skeleton structure as well as on the statistical results of the group-comparison patients versus controls is demonstrated in Fig. 7 (*in vivo dataset I*). In the first two rows, different numbers of repetitions were used to calculate the diffusion tensors. The influence of noise level decreases from left to right. The results show increasing numbers of false-positive tract centers in the skeleton structure with increasing noise levels (green arrows). Furthermore, it can be seen that significant group differences between patients and controls were detected even on those purely noise induced structures (blue arrows).

The level of significance for group differences between patients and controls was also heavily dependent on the noise level. The FA of the fornix, in this example, (red arrows) was significantly different between groups when using one repetition, but not when two repetitions were used. At three repetitions, the FA of the fornix again appeared as significantly different between patients and controls, while slightly decreasing in significance when going from three to ten repetitions (from $p=0.02$ to $p=0.04$).

While Fig. 7a shows only one representing subset for each noise level (1, 2, 3 and 10 rep.), Fig. 7b shows four possible subsets with two repetitions each (subset 1 with repetition 1 and 2; subset 2 with repetition 3 and 4; subset 3 with repetition 5 and 6; subset 4 with repetition 7 and 8). The FA skeleton differed only slightly from subset to subset. While subset 1 did not yield significant differences between patients and controls in the fornix, the remaining subsets did show significant

differences in this area. Apart from the fornix, similar effects were found in other areas of the skeleton (black arrows).

For *in vivo dataset II* similar results were obtained (Supplement 2). The 45 diffusion directions of this dataset were split into two subsets of 22 diffusion directions. The significance maps differed between noise levels (22 vs 45 diffusion directions) as well as between the subsets (22 vs 22 diffusion directions).

3.4 Influence of the user

Fig. 8 shows the different results obtained by using different registration targets. For *in vivo dataset I* with 30 subjects, the FMRIB58 target is characterized by a smoother FA map and a more clear depiction of the major WM structures in comparison to the study-specific registration target. This directly affected the structure of the FA skeleton (blue arrows) and the statistical results (green arrows). The fornix exhibited significant group differences when using the FMRIB58 target, but did not reveal significant group differences when using the study-specific atlas. This statement also holds true for *in vivo dataset II* with 100 subjects (green arrows), although differences in the mean FA skeleton shape due to the different registration targets are much less pronounced (blue arrows). For dataset II, additionally a study specific target is created with DTI-TK. The results obtained with this target are consistent with the results of the study-specific TBSS approach. One advantage of the DTI-TK study specific target creation is that the computation time scales with n, whereas it scales with n^2 for the TBSS approach. The standard TBSS procedure requires 10,000 pair-wise image registrations for n=100 subjects.

Fig. 9 shows the effect of varying the FA threshold in the skeletonization process. For *in vivo dataset I*, at a low threshold of 0.15, the FA skeleton also includes finer structures that disappear at higher threshold values. This also included some false-positive tract centers that we could not associate with any underlying WM tracts (see blue arrows, dataset I). At increasing threshold values, some known WM structures shrink or disappear (e.g. the capsula externa, fornix, cerebellum, see green arrows, dataset I/II). Interestingly, the significance levels between patients and controls were also altered for different threshold levels in both datasets. The fornix, for example, is present on all skeletons, but exhibited significant differences only at the lower threshold levels. The CC is characterized by a high FA (up to 0.9) and, here, neither the skeleton nor the statistical results are affected by the relatively small changes in the FA threshold (see red arrows, dataset I/II).

4. Discussion

TBSS is by far the most popular approach for performing voxel-wise DTI analyses. It provides dedicated processing steps and deals with smoothing and misalignment issues in diffusion MRI-based

group analysis studies. However, it also builds upon a certain set of assumptions that we have investigated in detail in this work. Most TBSS users are well-informed about the major processing steps and well-aware of their major weaknesses, such as the abandonment of directional information in the skeletonization process. Unfortunately, though, this knowledge is not of much use when interpreting the final results of a TBSS study. TBSS results usually do look very appealing, and it is impossible to quantify or even see any underlying ambiguities in the data without taking further efforts and looking deeper into the data. In fact, while some publications have discussed potential improvements or weaknesses of TBSS (Edden and Jones, 2011; Keihaninejad et al., 2012; Van Hecke et al., 2010; Zalesky, 2011), it is mostly unknown how much these weaknesses can actually impact the final results of a typical group comparison and the conclusions drawn from it. In the present study, we explore several key issues in this regard in order to further raise awareness of the pitfalls of TBSS and to provide constructive suggestions for future improvements of the technique.

Anatomical inaccuracies at the skeleton projection step

One of our major findings is the extent of anatomical inaccuracies that is inherent to the FA skeleton projection and the substantial bias that it can introduce. TBSS is known to be purely FA based, and it was previously reported that adjacent WM tracts are not necessarily separable based only on their FA (Kindlmann et al., 2007; Yushkevich et al., 2008). It was yet unknown, though, that the percentage of voxels that is misassigned to the wrong tract reaches such high numbers in two prominent tracts in the brain, the body of the CB and the CC. Interestingly, exactly this separation of the superior CB and the CC was explicitly stated to be solved and assumed to "work well" in the original TBSS publication (Smith et al. (2006), page 1494, second paragraph) despite the lack of any analysis in that article to substantiate this claim. We have shown that this assumption is not met and we have provided examples where inadequate separation of adjacent WM tracts occurs, even under ideal conditions, which are, perfect registration, no partial voluming, and infinite SNR. Under real conditions with residual misalignment, noise contributions, and partial volume effects, an even larger bias originating from the projection step can be expected.

Do these confounding factors really make a difference? In other words, should one worry about the validity of the outcome results given these issues, or could one simply proceed and assume that these effects are negligible? One could possibly argue that a decreased confidence in the projection step will only increase the variability and, thus, may just lower the sensitivity of the technique for finding potential changes between groups. However, as we have shown in this work, the situation is much more severe. With standard TBSS settings, the complete superior CB was incorrectly identified as being significantly different between groups, which was purely due to anatomically inaccurate assignments during the skeleton projection procedure. This finding was consistent in both datasets

that differ strongly considering number of subjects, field strength and number of gradient directions. Thus, our data suggests that this is an artefact that may occur regardless of the exact dataset description. There are two main factors that can contribute to the observation that the FA in a large region would appear significantly different between groups as a result of voxel misassignment. First, since the projection depends on the quality of each subject's alignment with the skeleton, which, in turn, is tightly bound to the morphology of the subject, the quality of assignments is highly group dependent if the disease at interest moderates morphology and not only microstructure (as demonstrated in the box plots in Fig. 2h). Second, due to the fact that TFCE accounts for statistical support from adjacent voxels in order to detect statistically significant differences in voxel clusters, the missing or increasing support of voxels in close proximity can quickly spread over the structure and can dramatically change the overall significance map. The statistical results obtained with TFCE are thus also influenced by the overall number of neighborhood voxels (i.e. the size of the skeleton sheet structure).

Bias at the skeleton projection step

Digging deeper into the skeleton projection step related to anatomical specificity, we have performed detailed analyses of the behavior of TBSS in images of a physical phantom with precisely defined fiber bundles with a diameter in the order of the voxel resolution. The assumed benefit of the TBSS projection step to compensate post registration alignment errors was previously analyzed by Zalesky (2011), in which it was reported that TBSS cannot compensate 90% of errors, but still gives good correspondence in the FA values. Looking at finer bundles such as the fornix, we expected that this FA value correspondence will also be strongly reduced. We were able to demonstrate that the positioning of the acquisition matrix and concomitant partial volume effects caused errors in the skeletonization projection, which is in line with previous findings by Edden and Jones (2011). One of the added benefits in this work is that we included well-defined phantom data that could act as a ground-truth reference of the underlying fiber architecture. One finding regarding the phantom experiments was particularly intriguing: we expected to find the highest FA values on the skeleton in cases where the imaging matrix is perfectly aligned with the phantom fiber tracts. In addition, with imperfect alignment the FA was expected to be lower due to partial voluming (Bach et al., 2013). However, when the TBSS skeleton projection was applied, the contrary was found. We have shown that this effect occurs due to the ill-defined skeleton sheet orientation and the related projection path. This effect is quite relevant, also when looking at in-vivo datasets, especially when analyzing finer tubular (e.g., the fornix) or circular (e.g., the uncinate fasciculus) structures. In general, the dominant factor that defines the orientation of the skeleton sheet may actually be related to the variation in anatomical alignment, rather than by the shape of the structure. In other words, the

smearing effect of imperfectly aligned structures when creating the mean FA template may lead to artificial sheet- or tubular-looking structures in the skeleton and can make a correct projection of the original structure impossible.

While many authors might not be aware of this effect, the authors of the TBSS publication have briefly noticed potential problems with small tubular structures like the fornix. In particular, in their results, they confirmed the quality of the projection vectors in the fornix to ensure that their result is not a finding based on pure chance. As such a confirmation step would be advisable for every TBSS study that investigates finer structures, it would be a valuable future extension of TBSS to further simplify this type of verification within the application. However, looking at our in-vivo experiments in the fornix, the FA skeleton orientation seems to be primarily determined by the inter-subject variability of the fornix position rather than by its shape, leading to a vast amount of voxels on the fornix skeleton that project directly into the cerebrospinal fluid and which do not belong to the fornix at all.

It is important to remember that in regions with complex fiber architecture, such as the area where the CC and corticospinal tract kiss/cross, it is much harder or even impossible to differentiate individual tracts while generating the skeleton and performing the projection procedure. In this context, we want to emphasize that the skeleton should be referred to as the FA skeleton, not the tract-skeleton, and that statistical results on this skeleton should be interpreted with these assignment problems in mind. A promising future improvement of TBSS could be to implement a skeletonization and projection step that does not ignore the directional information in the data. Yushkevich et al. (2008) use, for example, fiber tractography in order to distinguish between adjacent tracts. Until such a technique is developed in TBSS, one could consider an extended use of the “extra-treatment” that was originally added to manually guide the skeleton projection in the temporal cingulum as one of the important tubular structures in the brain (Smith et al., 2006). However, a clear distinction between tubular and sheet-like is not always possible and the required regions of interest would have to be drawn manually to produce a study-specific template. Unfortunately, this would be a major obstacle for the usability of TBSS and would also further reduce the objectivity of such analyses.

Statistical power and sensitivity to pathologies

The original TBSS publication provided insights into the repeatability of FA measurements across sessions and across subjects (Smith et al., 2006), reporting an inter-session coefficient of variation between 3% and 5%, and an inter-subject coefficient of variation of between 5% and 15%. These numbers, however, were derived by manually defining and comparing 7 voxels of interest on the skeleton for different major structures and did not include important aspects that come up when

considering the entire processing pipeline. In our experiments, we demonstrated the significant impact of noise on the final TBSS result. We showed that the noise level strongly affects the significance values of specific structures in the skeleton. We noticed that in terms of statistical significance these structures tended to appear or disappear as a whole rather than on voxel-level. We also observed this effect when varying the subsets for analysis while keeping the same level of noise and when changing the noise level for each subject's dataset. While such effects can dramatically change the conclusions drawn in a study (Bells et al., 2012), these may also be attributed in part to the TFCE approach (Smith and Nichols, 2009).

Another problem that we identified in our experiments is the noise-dependency of the shape of the FA skeleton. This is critical not only because TFCE depends on the skeleton shape, but also because we have identified significant group differences on artificial, noise induced structures that are anatomically meaningless (e.g., a skeleton part within the cerebro-spinal fluid). Thus, statistical correction is a major and important area of research in the future.

The current trend of increasing the resolution in diffusion weighted MRI potentially intensifies the problems of skeleton-based analyses. Reducing a full tract bundle to a one voxel thick skeleton becomes increasingly problematic with smaller voxel sizes (higher resolutions) with regard to statistical power, since a much higher percentage of the information in the image gets eliminated in the projection step. Therefore, and in light of increasingly accurate registration schemes and multi-compartment modeling, the original motivation of TBSS and the skeleton projection may need to be reconsidered. Note that this is in line with a current study that shows improved results via high-quality non-linear registration as compared to the registration-projection in TBSS (de Groot et al., 2013). This optimized registration approach is also sensitive to pathologies that may be overlooked using TBSS, e.g. in cases where the tract perimeter and not the tract center is affected by a disease. It is obvious that TBSS should not be used for topology changing diseases such as brain tumors.

TBSS is state of the art – some recommendations for use

Despite the methodological considerations presented in this work, TBSS is still the leading technique for voxel-wise DTI analyses at the moment as many alternative approaches are far less reproducible and may have similar problems in many of the discussed situations. In addition, one of the major strengths of TBSS is the minimal input required from the user. To encourage TBSS users to maximize the robustness and validity of their analyses we would like to conclude our discussion with suggestions for best practice. Two major obstacles for TBSS becoming completely objective are the degrees of freedom in the interpretation of results and the remaining parameter settings of the method.

First, the unambiguity in interpretation of the results is particularly unwarranted if studies only show a single arbitrary slice position from the multiplanar image maps. This problem could be alleviated by showing multiple equidistant slices in the image. Furthermore, when reporting and interpreting results, as demonstrated experimentally in this paper, this should be done with great care and, ideally, only after a check of the plausibility of the results. For example, for structures that are in close proximity to each other, such as the CC and CB, the potential influence of post-registration misalignments and voxel misassignments could be checked using a similar approach as we have adopted in our experiments, that is, by following the segmentations of structures throughout the pipeline. A further post-hoc evaluation can be performed by splitting the healthy controls into two groups and looking for any unexpected false-positives when performing TBSS on these two groups. In this design, no regions with significant differences are then expected. Furthermore, TBSS offers an "extra-treatment" to manually guide the skeleton projection for tubular structures. As we have shown that the standard projection procedure leads to significant artifacts in tubular structures, this extra-treatment should be considered whenever tubular structures are of special interest to a study.

The second obstacle, which is related to the parameter settings in TBSS, is much harder to tackle. The parameters allow the method to suit many different scenarios with different requirements on the one hand, but they leave room for tweaking TBSS to produce nice-looking results that are not really stable to obtain. The choice of the registration target, for example, has previously been investigated by (Keihaninejad et al. (2012)) and it was proposed to apply group-wise atlas construction in order to improve the alignment of DTI data and, consequently, the specificity and sensitivity of TBSS-results. In our analyses, we further investigated the effect of choosing different registration targets, and noticed that a large variation is introduced in the FA skeleton geometry and, subsequently, in the final statistical results. Keihaninejad et al. (2012) reported that the fornix appears significantly different between AD patients and controls when registering to the FMRIB58 template and that the significance vanishes if a group-wise atlas is chosen as registration target. We analyzed this using two further AD datasets with up to one hundred subjects and the effect emerged even more clearly. In contrast to Keihaninejad et al. (2012) differences in the statistical results already occurred when switching between the two TBSS standard options: 1. registration to the FMRIB58 template or 2. registration to the most representative subject of the group. Similar findings were obtained by varying another important user setting, the FA threshold in the skeletonizing process. Again, some structures were detected to be significantly different between groups for one setting, but not for the other. This is precarious, since many users do not have the knowledge or expertise to evaluate such effects in detail. In addition, the optimal parameters for their specific study cannot be known in advance.

As a consequence, we propose a clear rule for TBSS studies in this regard: only report results that are based on the default parameter settings given by TBSS, as long as there is no clear evidence from literature not to do so. All settings that deviate from the default configuration in TBSS should be explicitly mentioned and motivated. In addition, the stability of findings with regard to the FA skeleton threshold should be checked for low-FA structures like the fornix or the capsula externa as these can be particularly unreliable. Furthermore, with regard to the choice of the registration target, a recommendation can already be made on basis of both previous studies and our work: replace the TBSS registration step with tensor-based, group-wise registration, e.g. using DTI-TK.
Our recommendations for a specific TBSS processing pipeline are summarized in table 1. Two further tables show our recommendations regarding the interpretation of TBSS results (table 2) and future improvements of TBSS (table 3). Finally, to ensure complete reproducibility and examination of the results, we encourage researchers to make their datasets available (either publicly or upon request), which is already common practice in many other research fields.

Acknowledgements: This study was partly funded by the German Research Counsel (DFG, LA 2804/1-1) and the contribution of Chantal Tax is supported by an FC-EW grant (No. 612.001.104) from the Dutch scientific foundation (NWO). The authors would like to thank the members of the Utrecht Vascular Cognitive Impairment Study Group for providing the diffusion MRI data (dataset II): University Medical Center Utrecht, the Netherlands, Department of Neurology: E. van den Berg, G.J. Biessels, M. Brundel, W.H. Bouvy, S.M. Heringa, L.J. Kappelle, Y.D. Reijmer; Department of Radiology/Image Sciences Institute: J. de Bresser, H.J.Kuijf, A. Leemans, P.R. Luijten, W.P.Th.M.Mali, M.A. Viergever, K.L. Vincken, J.J.M. Zwanenburg; Department of Geriatrics: H.L. Koek, J.E. de Wit; Hospital Diakonessenhuis Zeist, the Netherlands: M. Hamaker, R. Faaij, M. Pleizier, E. Vriens.

References

Bach, M., Fritzsche, K.H., Stieltjes, B., Laun, F.B., 2013. Investigation of resolution effects using a specialized diffusion tensor phantom. Magn Reson Med.

Bells, S., Dustan, L., McGonigle, D.J., Evans, C.J., Jones, D.K., 2012. On the Stability of Skeleton-Based Analyses of Diffusion Tensor MRI-based Measures. Proc. Intl. Soc. Mag. Reson. Med. 20, Melbourne.

de Groot, M., Vernooij, M.W., Klein, S., Ikram, M.A., Vos, F.M., Smith, S.M., Niessen, W.J., Andersson, J.L., 2013. Improving alignment in Tract-based spatial statistics: Evaluation and optimization of image registration. Neuroimage 76, 400-411.

Edden, R.A., Jones, D.K., 2011. Spatial and orientational heterogeneity in the statistical sensitivity of skeleton-based analyses of diffusion tensor MR imaging data. J Neurosci Methods 201, 213-219.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63-72.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825-841.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62, 782-790.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med Image Anal 5, 143-156.

Jones, D.K., Symms, M.R., Cercignani, M., Howard, R.J., 2005. The effect of filter size on VBM analyses of DT-MRI data. Neuroimage 26, 546-554.

Keihaninejad, S., Ryan, N.S., Malone, I.B., Modat, M., Cash, D., Ridgway, G.R., Zhang, H., Fox, N.C., Ourselin, S., 2012. The importance of group-wise registration in tract based spatial statistics study of neurodegeneration: a simulation study in Alzheimer's disease. PLoS One 7, e45996.

Kindlmann, G., Tricoche, X., Westin, C.F., 2007. Delineating white matter structure in diffusion tensor MRI with anisotropy creases. Med Image Anal 11, 492-502.

Reijmer, Y.D., Leemans, A., Caeyenberghs, K., Heringa, S.M., Koek, H.L., Biessels, G.J., Utrecht Vascular Cognitive Impairment Study, G., 2013. Disruption of cerebral networks and cognitive impairment in Alzheimer disease. Neurology 80, 1370-1377.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum Brain Mapp 17, 143-155.

Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487-1505.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83-98.

Tournier, J.D., Mori, S., Leemans, A., 2011. Diffusion tensor imaging and beyond. Magn Reson Med 65, 1532-1556.

Van Hecke, W., Leemans, A., D'Agostino, E., De Backer, S., Vandervliet, E., Parizel, P.M., Sijbers, J., 2007. Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information. IEEE Trans Med Imaging 26, 1598-1612.

Van Hecke, W., Leemans, A., De Backer, S., Jeurissen, B., Parizel, P.M., Sijbers, J., 2010. Comparing isotropic and anisotropic smoothing for voxel-based DTI analyses: A simulation study. Hum Brain Mapp 31, 98-114.

Van Hecke, W., Leemans, A., Sage, C.A., Emsell, L., Veraart, J., Sijbers, J., Sunaert, S., Parizel, P.M., 2011. The effect of template selection on diffusion tensor voxel-based analysis results. Neuroimage 55, 566-573.

Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., Jeurissen, B., 2013. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. Neuroimage 81, 335-346.

Wang, Y., Gupta, A., Liu, Z., Zhang, H., Escolar, M.L., Gilmore, J.H., Gouttard, S., Fillard, P., Maltbie, E., Gerig, G., Styner, M., 2011. DTI registration in atlas based fiber analysis of infantile Krabbe disease. Neuroimage 55, 1577-1586.

Yushkevich, P.A., Zhang, H., Simon, T.J., Gee, J.C., 2008. Structure-specific statistical mapping of white matter tracts. Neuroimage 41, 448-461.

Zalesky, A., 2011. Moderating registration misalignment in voxelwise comparisons of DTI data: a performance evaluation of skeleton projection. Magn Reson Imaging 29, 111-125.

Zhang, H., Avants, B.B., Yushkevich, P.A., Woo, J.H., Wang, S., McCluskey, L.F., Elman, L.B., Melhem, E.R., Gee, J.C., 2007. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. IEEE Trans Med Imaging 26, 1585-1597.

Zhang, H., Awate, S.P., Das, S.R., Woo, J.H., Melhem, E.R., Gee, J.C., Yushkevich, P.A., 2010a. A tract-specific framework for white matter morphometry combining macroscopic and microscopic tract features. Med Image Anal 14, 666-673.

Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C., 2006. Deformable registration of diffusion tensor MR images with explicit orientation optimization. Med Image Anal 10, 764-785.

Zhang, H., Yushkevich, P.A., Rueckert, D., Gee, J.C., 2010b. A computational white matter atlas for aging with surface-based representation of fasciculi. Biomedical Image Registration. Springer, pp. 83-90.

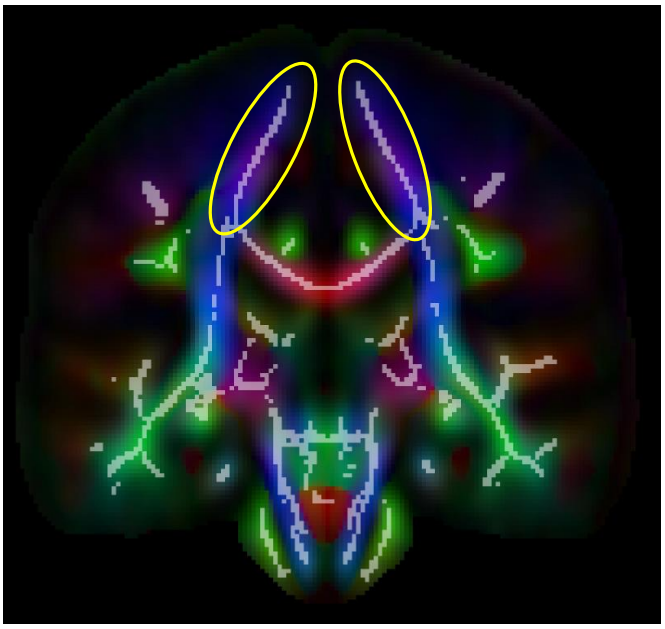

Fig. 1 – Collapse of different white matter tracts. In regions where pathways of different structures merge (yellow ellipses), the skeletonization step causes these different bundles to collapse on top of each other. Therefore, it is virtually impossible to assign the FA values to the same anatomical structure across subject in a consistent way. **Legend**: colored – tensor color map of a human brain; white – FA skeleton; yellow ellipses – regions where the superior part of the CC and the corona radiata merge.

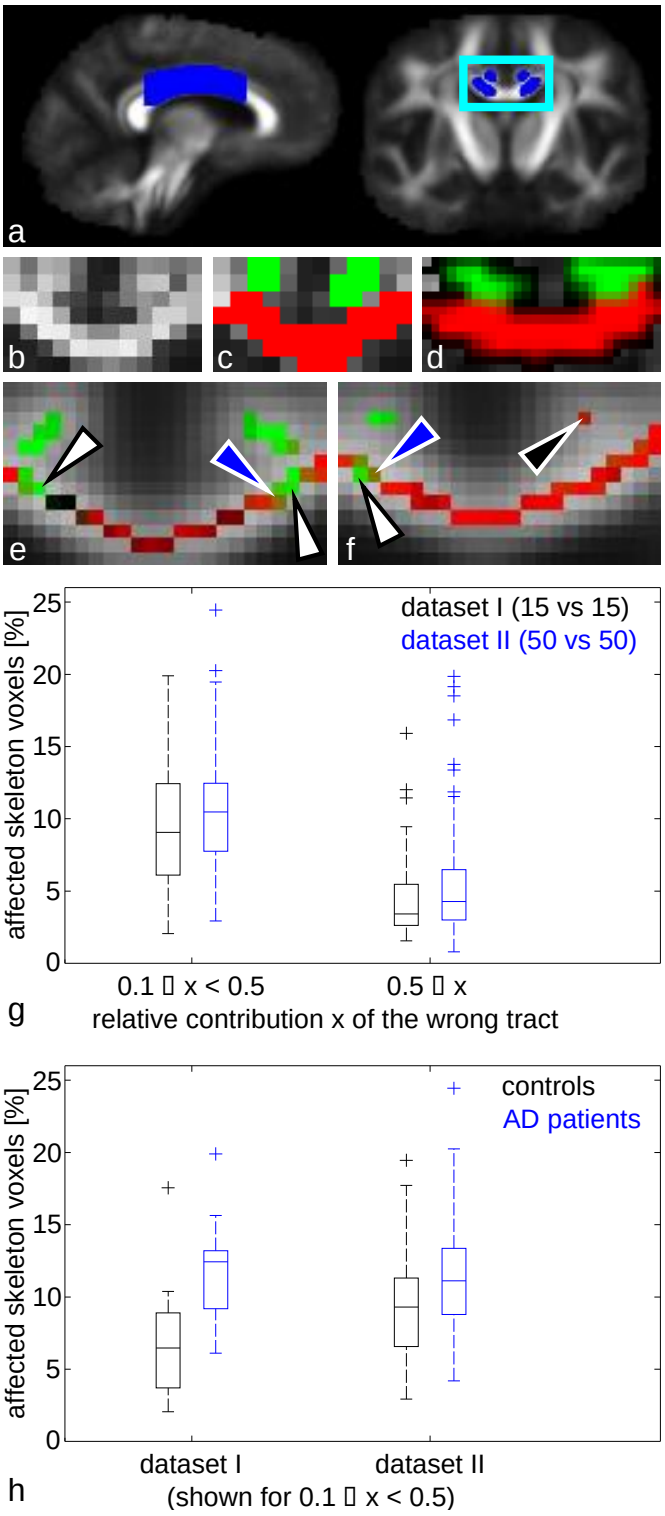

Fig. 2 – Influence of adjacent white matter tracts. (a) Sagittal and coronal view of an exemplary FA map and the region of interest (blue). The cyan highlighted area is shown in b-f. **(b)** Coronal view of FA values in the region of interest. **(c)** Segmentation of the CB (green) and the CC (red). **(d)** Segmentation after registration and resampling. **(e+f)** Result of the skeleton projection step in two different subjects. Some CB-voxels were assigned to the CC-skeleton (white arrows) and vice versa (black arrow). Blue arrows indicated voxels whose FA values are a mixture of CB and CC FA values.

Black skeleton voxels represents voxels, which could not be identified as CB or CC voxels previously. **(g)** Fraction of voxels per subject that had a relative contribution x of the "wrong" tract to the FA values of the skeleton (*in vivio datasets I/II*). **(h)** Patients and controls differed significantly for 0.1≤x< 0.5 (p=0.001 for dataset I, p=0.039 for dataset II). Outliers are indicated by '+'.

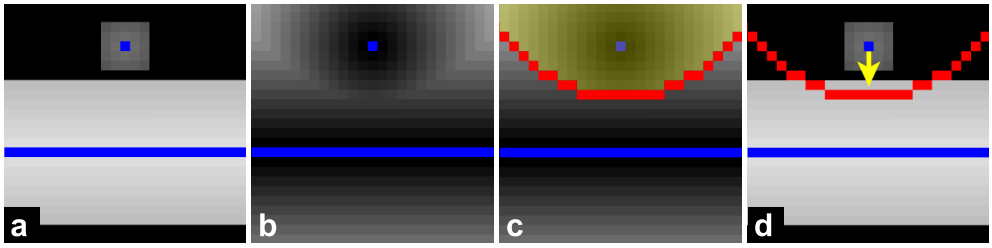

Fig. 3 – Potential source of misassignment. (a) Synthetic FA map (simulating a coronal view of the CC and the CB as in Fig. 2b). The obtained FA skeleton is shown in blue. **(b)** Distance map calculated by TBSS (higher intensities reflect larger distances to the skeleton). **(c)** The search area for local FA maxima for the upper fiber strand (yellow area). **(d)** The misassignment of a voxel from the bottom tract (the maximal FA value in the search area) to the upper skeleton (yellow arrow).

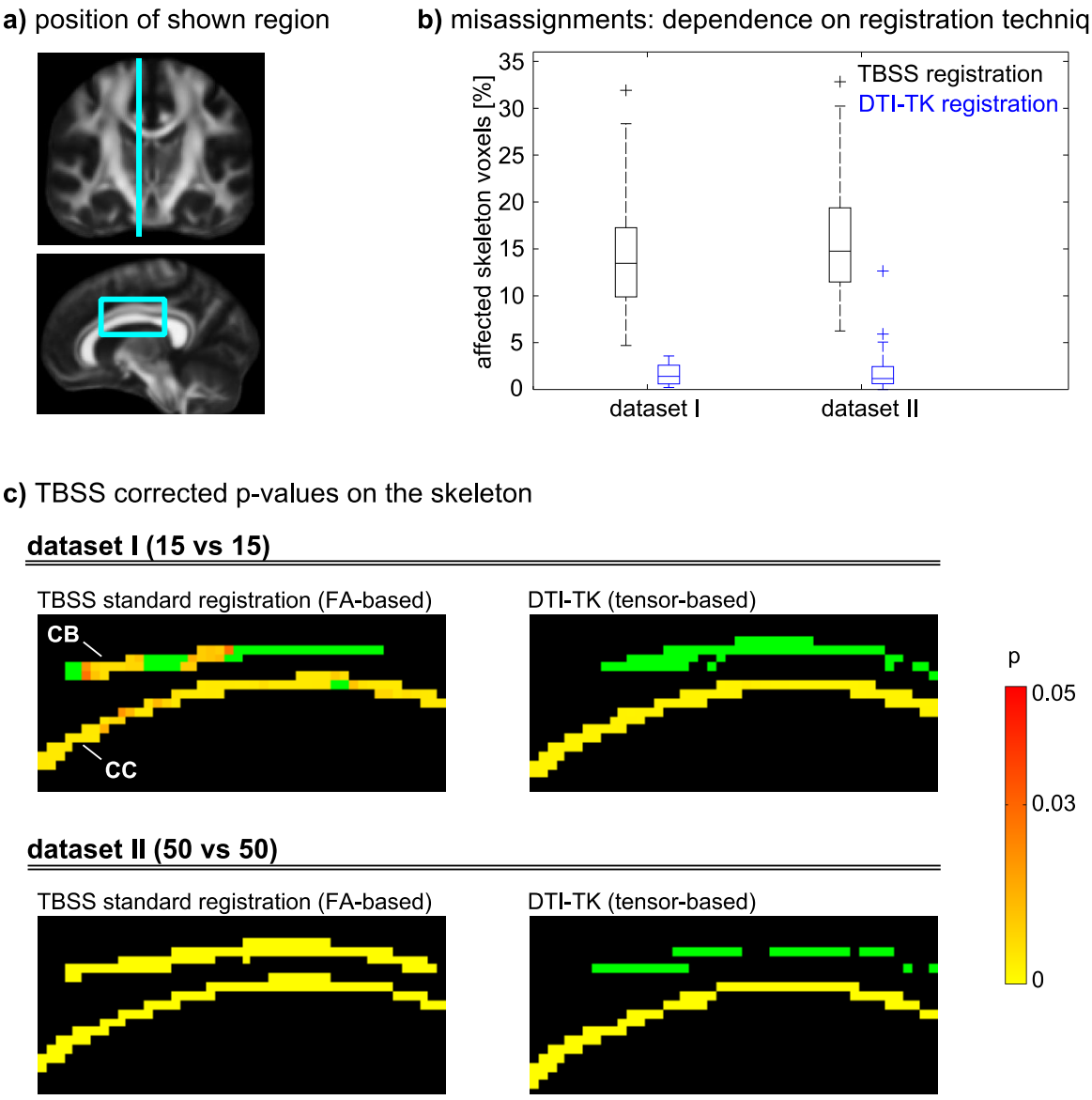

Fig. 4 – Impact of misassignments on TBSS results. (a) The position of the region shown in c is highlighted cyan colored. **(b)** The fraction of voxels per subject, which have a contribution of at least 10% of the neighboring strand, strongly depends on the registration technique. **(c)** Sagittal view (upper tract: CB, lower tract: CC) of the TBSS statistical results obtained by the TBSS standard pipeline, as well as with an advanced tensor-based registration technique (DTI-TK). Both in vivo dataset types are investigated. The DTI-TK registration decreased the number of misassigned voxels by factor seven. The highly significant differences in the CB that were identified by the conventional pipeline completely disappear when the number of misassignments is decreased.

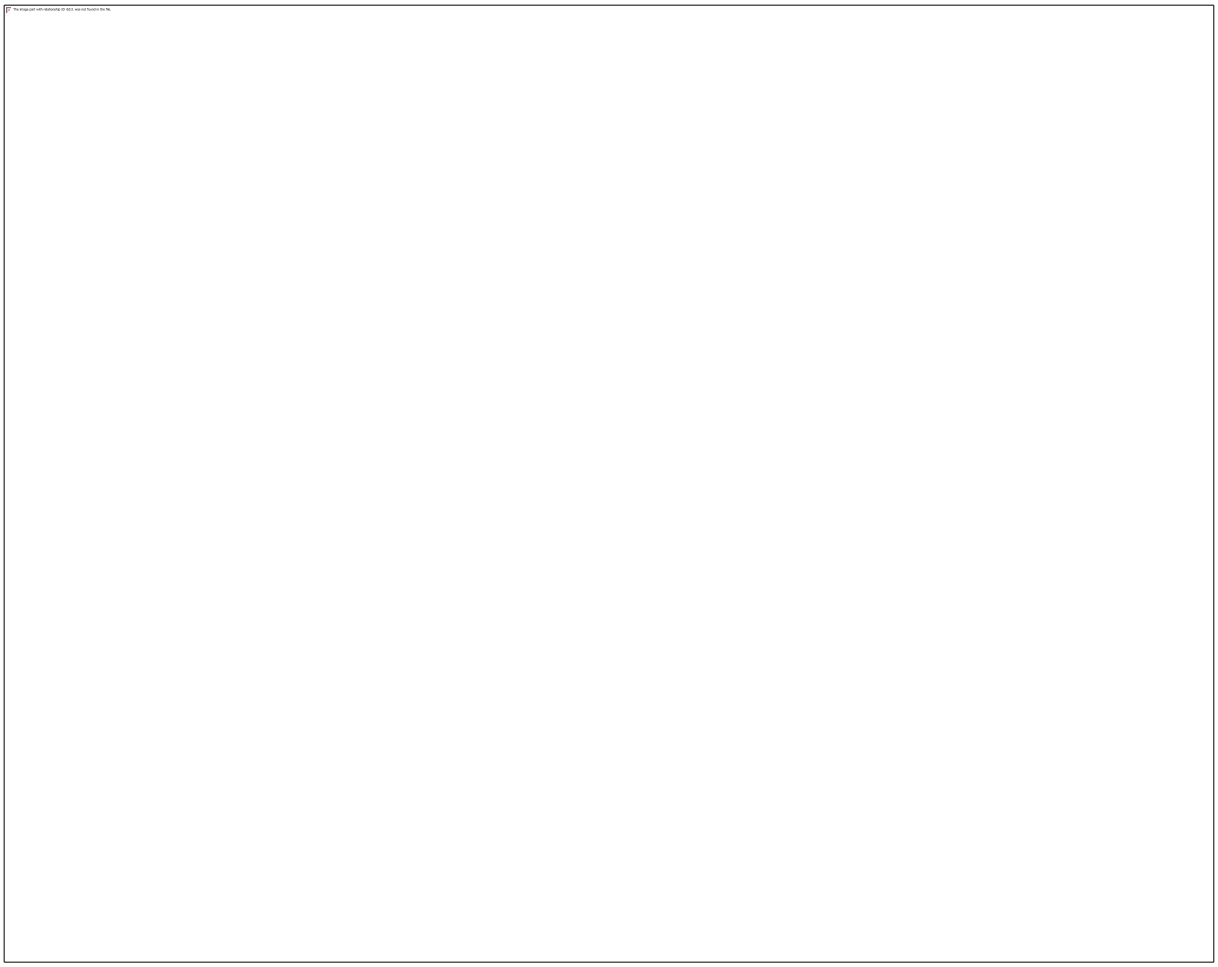

Fig. 5 – mean FA skeleton and anatomical concordance in individual subjects. (a) Because of the mismatch between search and misalignment direction, the TBSS skeletonization step could not compensate residual registration misalignments in the fornix. The area highlighted in cyan is depicted in b and c. **(b+c)** Midsagittal view of the fornix with good (green arrows) and bad (red arrows) concordance between the mean FA skeleton and the fornix. Five subjects are shown for each in vivo dataset type. The red arrows indicate areas where the FA skeleton does not cover the fornix and the search for the tract center has to fail. **Legend:** transparent blue / blue – mean FA skeleton; grayscale background – FA maps.

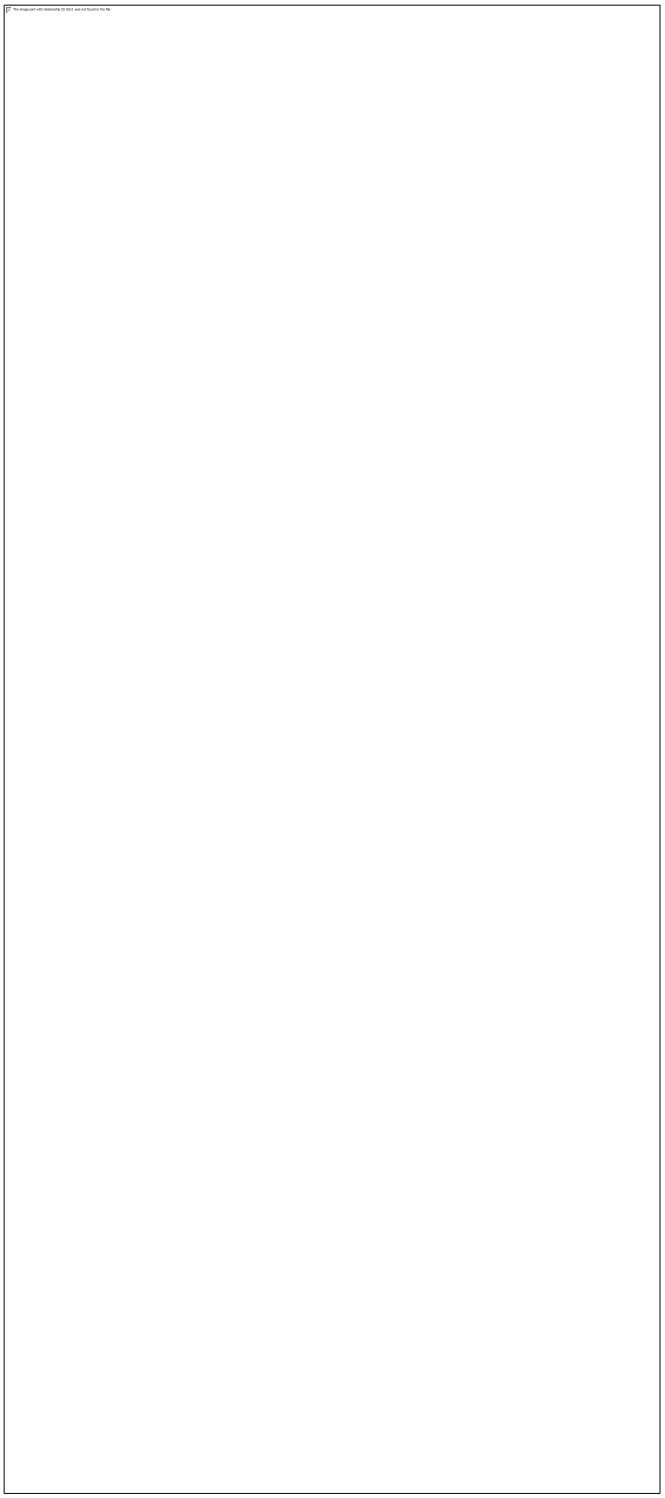

Fig. 6 – Influence of resolution/partial volume and skeleton shape. (a) Circular phantom spindle with 6 fiber strands (blue) of square cross-section. **(b)** Slice through the phantom FA image using the FOV that was illustrated in (c) by the green (b, right) and red (b, left) squares. **(c)** Schematic depiction of one of the fiber strands and two different positions of the FOV (green and red squares). **(d)** Side view (top) and top view (bottom) of the FA skeleton (green) of the 2.5 mm thick strand for the FOV that was illustrated in green. **(e)** Same as (d), but using the FOV that was illustrated in red. Please note the flipping orientation of the FA skeleton sheets in green vs. red, resulting in different search

directions for the TBSS projection step. **(f)** The frequency distribution of the projected FA values for the red and the green skeleton. The FA values on the red skeleton tended to be lower, even though the red configuration was much less effected by partial volume effects than the green configuration.

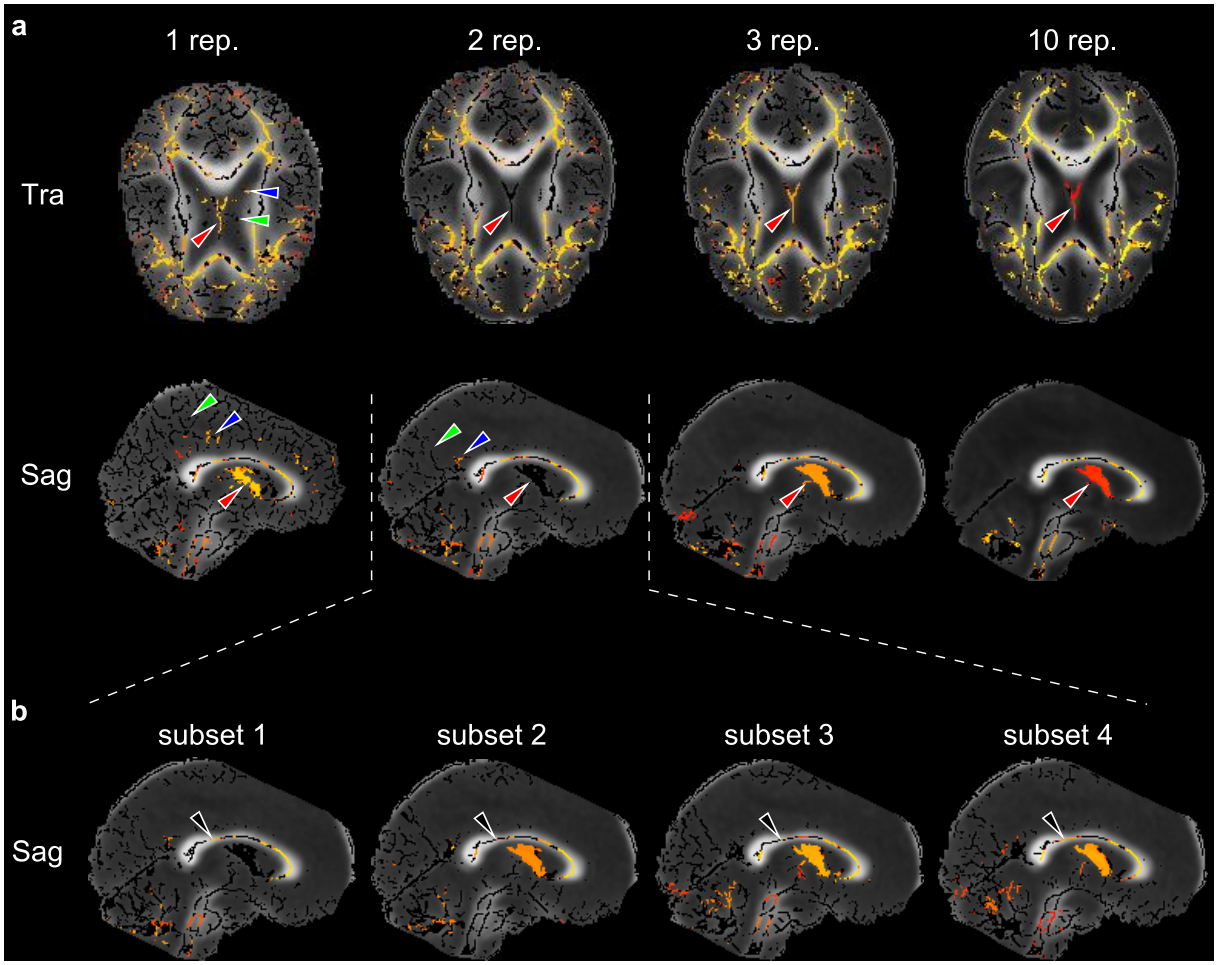

Fig. 7 - Influence of image noise. **(a)** TBSS significance maps obtained using different numbers of repetitions (1, 2, 3 and 10, *in vivo dataset I*). Higher noise levels lead to false-positive tracts in the skeleton (green arrows). Some false-positive tracts were subject to significant group differences (blue arrows). The noise induced parts of the skeleton largely disappear when using all 10 repetitions. The significance levels of group differences (e.g. in the fornix) go up and down for different numbers of repetitions (red arrows). (**b**) Four possible subsets of two repetitions each are shown. The FA skeleton differs only slightly from subset to subset. The fornix shows significant differences in 3 of the 4 subsets. Significance of differences also changes in the CC (black arrows). Corresponding results were obtained for *in vivo dataset II* (see Supplement 2). **Legend**: grayscale background – mean FA; black lines – FA skeleton; colors ranging from red to yellow – significant ($p \leq 0.05$) differences between Alzheimer's disease patients and controls (same colorbar as in Fig. 4, red: low significance, yellow: high significance); Tra – transversal view; Sag – sagittal view.

registration target
mean FA
statistical results
Tra
Sag
dataset I (15 vs 15)
FMRIB58
study specific FSL
dataset II (50 vs 50)
FMRIB58
study specific FSL
study specific DTI-TK

Fig. 8 – Influence of the user I: registration target. (legend similar to Fig. 7) TBSS results were generated using the FMRIB58 target and then compared to TBSS results obtained by using a study-specific target. For dataset I the study-specific target exhibited a more brittle FA skeleton with potentially false-positive tract centers (blue arrows). This effect vanishes for dataset II with 100 subjects. The statistical results were also influenced by the choice of the registration target (see green arrows, e.g. in the fornix). This statement holds true in both datasets. For dataset II an alternative study specific target is created with DTI-TK and the results are consistent with the study specific TBSS approach.

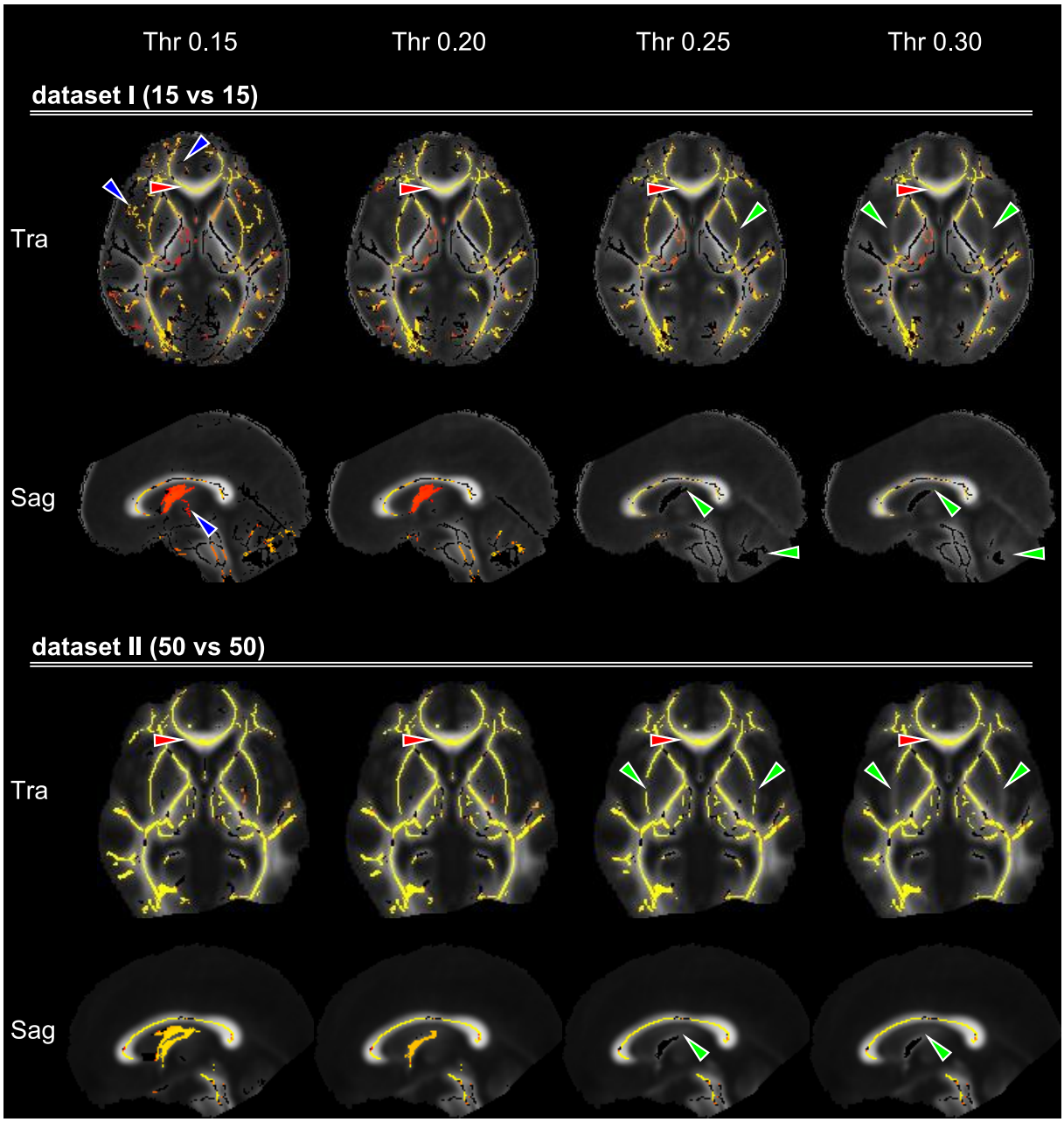

Fig. 9 – Influence of the user II: FA threshold. (legend similar to Fig. 7) TBSS was performed at different FA thresholds. For dataset I the FA skeleton at lower FA thresholds shows finer but potentially false-positive white matter tracts (blue arrows). At higher thresholds, the white matter

tracts are more precisely defined. These effects vanish if a higher number of subjects is used (dataset II). The significance level of group differences was highly dependent on the FA threshold for both datasets. The fornix, for example, was represented on all skeletons, but was only found to be significantly altered at threshold levels of 0.2 or below. At higher thresholds, important structures begin to disappear from the skeleton (e.g. the capsula externa, fornix, cerebellum, see green arrows). The FA skeleton as well as statistical results at the CC are not influenced by the relatively low changes in the FA threshold (see red arrows).