An Analysis of Variability in Diffusion Tractography of Language Fascicles

Kesshi Marin Jordan[1], Eduardo Caverzasi[2,3], Valentina Panara[1,4], Bagrat Amirbekian[1], Anisha Keshavan[1], Nico Papinutto[2,5], Mitchel Berger[6], and Roland Henry[2]
[1]Bioengineering, University of California San Francisco & Berkeley, San Francisco, CA, United States, [2]Neurology, University of California San Francisco, San Francisco, CA, United States, [3]University of Toronto, Toronto, Canada, [4]Institute of Advanced Biomedical Technologies, University "G. D'Annunzio", Chieti, Italy, [5]Bioengineering, University of California San Francisco & Berkeley, CA, United States, [6]Neurosurgery, University of California San Francisco, San Francisco, United States

Target Audience – Researchers and surgeons that use probabilistic diffusion tractography to map white matter fascicles.

Purpose: Tractography methods are highly subject to implementation decisions such as algorithm choice, stopping criteria, initialization parameters, and threshold for defining fascicle volume (#streamlines/voxel). At present, presurgical methods to isolate a particular fascicle are highly dependent on a human operator, which introduces additional potential for variability. At our institution, we have implemented q-ball residual bootstrap probabilistic tractography in pre-surgical planning for language pathways in brain tumor patients. We were able to apply this novel technique consistently in a large cohort of tumor patients (more than 300 since January 2012), but characterization of variability is necessary. This study investigates inter- and intra- operator variability of fascicle volume definition in both control subjects and tumor patients, as well as how these values vary as a function of threshold. The subsequent effects of threshold on the calculation of a fascicle's average diffusion metrics was also investigated to evaluate the potential bias introduced by failure to characterize this variability.

Methods: High-angular resolution diffusion imaging (HARDI) was performed on ten healthy controls (HC) and ten patients with high-grade gliomas (HGG) both pre- and post-surgery. Several language fascicles were reconstructed using q-ball probabilistic fiber-tracking in the left hemispheres by two independent operators, according to the methods reported in Caverzasi, et al.[1,2] (http://nipy.org/dipy/). The results were visualized using Trackvis (http://trackvis.org). Language fascicles investigated in all subjects included: the arcuate (AF), the inferior fronto-occipital (IFOF), the superior longitudinal (branches II & III: SLFip, and temporal-parietal components: SLFtp), and the uncinate (UF) fasciculi. The inferior longitudinal (ILF) and middle longitudinal (MdLF) fasciculi were also reconstructed in tumor patients.

For each reconstructed pathway, the volume of the binary fascicle mask was defined using a #streamlines/voxel threshold. The percent common voxel agreement (PCVA) of the binary mask results defined as the sum of the volumes of binary masks divided by the logical AND of the two masks (scaled by 200). The binary masks defining fascicle volume were based on several thresholds of the original density mask (>0, >1, >5, >10, >25, >50, >75 streamlines). These binary masks were registered to MNI space to compare the spatial concordance of track definition intra- and inter-operator across subjects. Diffusion metrics (Fractional Anisotropy: FA, Radial Diffusivity: RD, Mean Diffusivity: MD, and Primary Eigenvalue: L1) were also calculated for each track volume across operators and across thresholds. The intra-rater reliability of the IFOF in controls was also investigated.

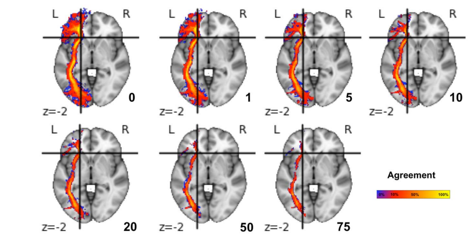

Figure 1: Spatial concordance of Intra-rater reliability of the IFOF in controls with increasing thresholds.

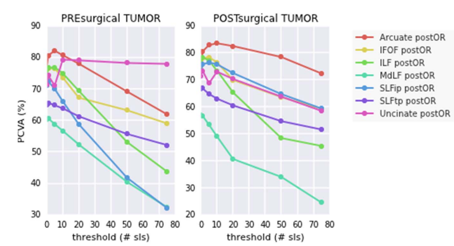

Figure 2: Inter-rater PCVA of language fascicles in the left hemisphere of tumor patients as a function of threshold

Results: These results indicate that probabilistic tractography methods tend to have an optimal threshold for maximum PCVA. Well-characterized fascicles, such as the AF and IFOF, show a peak around thresholds of 5 to 10 (Figure 2). This pattern is much more prominent in control subjects (data not shown). The spatial concordance analysis (Figure 1) demonstrates that a threshold of 5 largely eliminates disagreement voxels introduced by sparse streamlines at the periphery. After this point, however, the higher thresholds cut somewhat arbitrarily into the trunk of the fascicle volume. In the IFOF, intra-rater reliability is the highest (max ~84%) at a threshold of 5. If the same tracking instance is used (no variability due to probabilistic method), the pure intra-rater reliability is the highest (max~93%) with threshold of 0.

The choice of threshold in defining a track volume can bias the resulting average diffusion metrics, especially fractional anisotropy (Figure 3). The change of threshold from 0 to 1, for example, results in a 5% difference in average FA (across all tumor subjects and fascicles). The change from a threshold of 5 to 10 and 10 to 20, however, only results in a mean FA difference of about 3%. This suggests that for this particular tractography methodology, a threshold around 10 produces a relatively stable fractional anisotropy measurement.

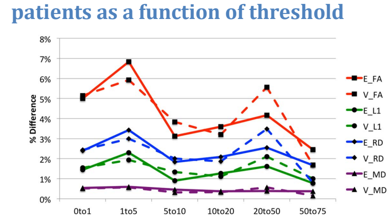

Figure 3: Effect of threshold change on average diffusion metrics in tumor subjects (line style: operator)

Discussion: The wide variety of inter-rater reliability is probably due to a combination of anatomy complexity and how easily the fascicle can be reconstructed. The better-documented fascicles, such as the AF, IFOF, and SLFip, tend to have higher PCVA values, while fascicles that are still very much under development, such as the MdLF, have much lower PCVA values.

The analysis of average diffusion metrics show striking average FA changes between commonly used thresholds. This FA variability is of a magnitude that has been reported as significant effect sizes in studies identifying average FA differences in fascicles. This suggests that studies employing these tractography methods for defining average fractional anisotropy should interpret their results in the context of a threshold-dependence analysis.

Conclusion: Diffusion tractography remains the only method of mapping the white matter of the brain noninvasively. As a result, this technology has been increasingly employed as a pre-surgical mapping tool for the motor, visual, and language systems. These methods have also spread to other fields as a means of defining a fascicle volume over which metrics are calculated and correlated with various parameters. In order for these technologies to be safely employed clinically, variability must be characterized and minimized for each tractography methodology. The use of diffusion metrics to describe fascicles defined by tractography has also been gaining popularity in fields outside of imaging. These metrics are particularly dependent on implementation choices, such as threshold, as the tractography itself depends on them as well. The characterization and minimization of these potential biases is important to ensure that tractography tools are used in a meaningful way.

References

1. Mandelli M, et al. Frontal white matter tracts sustaining speech production in primary progressive aphasia. J Neurosci. 2014; 34(29):9754-67.
2. Caverzasi E, et al. Identifying pre-operative language tracts and predicting post-operative recovery of function using q-ball HARDI diffusion MRI Fiber Tractography in patients with Gliomas. (submitted)