Quantitative comparison of automatic and manual tract segmentation methods

S. Muñoz Maniega[1], J. D. Bridson[2], W. J. Ang[2], P. A. Armitage[1], C. Murray[3], A. J. Gow[3], M. E. Bastin[4], I. J. Deary[3], and J. M. Wardlaw[1]
[1]Clinical Neurosciences, University of Edinburgh, Edinburgh, United Kingdom, [2]Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, United Kingdom, [3]Psychology, University of Edinburgh, Edinburgh, United Kingdom, [4]Medical Physics, University of Edinburgh, Edinburgh, United Kingdom

Introduction

Quantitative diffusion MRI tractography is widely used to study changes in white matter (WM) structure in a variety of neurological disorders. One of the main applications of tractography is in its use as a segmentation tool, for example in group studies of WM, in which the same tract needs to be segmented reliably from subject to subject across the population. A novel method for automatic tract segmentation known as probabilistic neighbourhood tractography (PNT) produces reliable tract segmentation, with reproducibility and repeatability rates similar to those reported in studies applying well accepted tractography methods using manual seed placement and multiple region-of-interest (ROI) constraints [1]. Both methods can also yield similar segmentations of tracts; however the level of agreement between them has yet to be tested quantitatively. In this work we directly compare PNT and ROI tractography methods in the same dataset to study their level of agreement in terms of the averaged fractional anisotropy (FA) and mean diffusivity (MD) measured in a set of tracts.

Methods

Diffusion MRI data from 53 healthy older participants from the Lothian Birth Cohort 1936 [2] were used for this study (27 male; mean age 71.4 ± 0.5 years). Data were acquired using a 1.5T GE MRI scanner, and comprised diffusion-weighted images (b=1000 s/mm^2) acquired in 64 non-collinear directions, and 7 T_2-weighted images. The resolution was 2×2×2 mm^3. Data were pre-processed and diffusion tensor parameters estimated using FSL tools [3]. In both tractography modalities we used BEDPOSTx/ProbTrackx [4] with a two-fibre model and 5000 streamlines to estimate the tracts of interest. The tracts segmented were corpus callosum genu and splenium, left anterior cingulum bundle (lcing) and left arcuate (larc) and uncinate (lunc) fasciculi. ROI tract segmentation was performed by placing a seed ROI in the centre of the tract on a T_2-weighted map of the subject's brain; the size of the seed ROI in voxels was 2×2×2 for lcing, lunc and larc and 3×3×3 for genu and splenium. To remove subsets of false positive connections and achieve tract segmentations comparable with those obtained in PNT a 'not' operation was applied with manually placed exclusion masks [5]. An example of seed and exclusion ROI placement is shown in Figure 1. PNT was run with as many seed points selected as the equivalent number in the seeding ROI (i.e. 27 for corpus callosum and 8 for association tracts), with a minimum matching probability to the reference tract of 0.01 [6]. Average values of FA and MD were calculated for the tracts segmented with both tractography methods. Agreement of the two methods was tested using Pearson's correlations and Bland-Altman analysis of percentage differences [7]. ROI tractography was repeated by a different operator in a subsample of the subjects (n=27) to assess reproducibility.

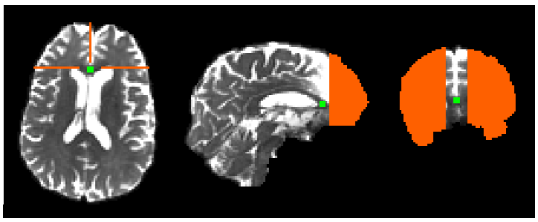

Figure 1: Seeding (green) and exclusion (orange) ROI for segmentation of the genu of the corpus callosum

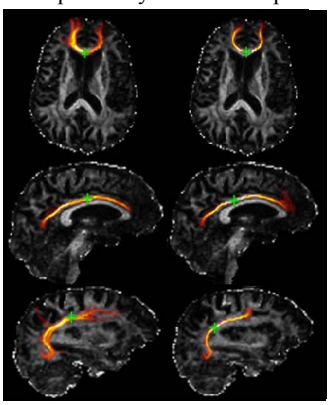

Figure 2: Projections of genu, lcing and larc obtained with ROI tractography (left) and PNT (right). Seed shown in green.

Results

Figure 2 shows examples of tracts segmented in one subject with both methods displayed with a standard threshold of 1% of the maximum connection probability. Visual inspection revealed that PNT tracts were generally more constrained within the main WM than tracts calculated from ROI.

FA was highly correlated between methods in genu, larc and lcing (r=0.60 to 0.80; p<0.001), with slightly weaker correlations for the splenium and lunc (r=0.38, 0.41; p=0.005, 0.004 respectively). Similarly, MD was highly correlated in genu, lcing and lunc (r=0.60 to 0.80; p<0.001), with medium correlations for splenium and larc (r=0.30, 0.40; p=0.031, 0.004 respectively).

% mean difference (LLoA, HLoA) **(PNT-ROI)**	FA	MD	% mean difference (LLoA, HLoA) **(Op1-Op2)**	FA	MD
genu	0.2 (-15.4, 15.8)	-2.0 (-16.3, 12.3)	*genu*	-0.3 (-10.5, 9.9)	-0.9 (-9.4, 7.5)
splenium	-9.0 (-30.8, 12.7)	16.2 (-9.3, 41.7)	*splenium*	-5.6 (-20.4, 9.2)	0.1 (-10.8, 11.0)
larc	5.2 (-9.6, 20.0)	-3.2 (-17.0, 10.6)	*larc*	-0.7 (-11.9, 10.5)	0.4 (-10.0, 10.7)
lcing	-4.9 (-25.0, 15.2)	6.0 (-6.2, 18.2)	*lcing*	-2.7 (-20.5, 15.1)	0.6 (-10.4, 11.6)
lunc	10.5 (-12.4, 33.4)	-2.6 (-16.6, 11.3)	*lunc*	3.4 (-12.4, 19.1)	-2.0 (-14.8, 10.8)

Op = operator; LLoA = Lower Limit of Agreement; HLoA = Higher Limit of Agreement

The table shows the results from the Bland-Altman analysis. Mean differences between the methods were ≤10% in FA and MD in all tracts except for the splenium MD. Reproducibility values (Op1 vs Op2) also show good agreement with mean differences <6%.

Discussion

The PNT and ROI methods differ greatly in their requirement for user input, as PNT requires very little operator input while ROI requires extensive manual ROI placement. Both methods show a reasonable agreement with mean differences of 10% being comparable to the reproducibility obtained when ROI are manually placed by different operators. The different approaches are, however, reflected in the 95% limits of agreement (LLoA, HLoA) which show greater variability in some tracts. This might be caused by the seeding method in PNT; even though an equivalent number of seed points was allowed in PNT and ROI analyses, only those with a minimum matching probability to the reference tract were kept in PNT, maintaining a high confidence in the segmentation. The splenium of the corpus callosum showed the weakest correlations and the lowest reproducibility for both within- and between- method comparisons; likely due to the large cross-section of this fibre where each attempt to segment this tract could be extracting a slightly different part of the fibre. In conclusion, PNT segmentation showed reasonable agreement with conventional ROI tract segmentation, with the advantage that given the same reference tracts and standard seed points the output segmentations for PNT will always be identical, removing operator dependency.

[1] Clayden JD et al 2009, NeuroImage 45:377-385; [2] Deary IJ, et al 2007, BMC Geriatr 7:28; [3] FMRIB, Oxford, UK, www.fmrib.ox.ac.uk; [4] Behrens TE et al 2007, NeuroImage 34:144-155; [5] Wakana S et al 2004, Radiology 230:77-87 ; [6] Clayden JD et al 2008, 16th ISMRM Meeting; p1840; [7] Bland JM and Altman DG 1999, Stat Methods Med Res 8:135-160