Tractography in the presence of multiple sclerosis lesions: Supplementary material

Ilona Lipp[a,b,c,1,*], Greg D Parker[b,d,1], Emma C Tallantyre[a,e], Alex Goodall[b,f,g], Steluta Grama[b], Eleonora Patitucci[b], Phoebe Heveron[a], Valentina Tomassini[a,b,e,h,2], Derek K Jones[b,i,2]

[a]*Division of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK*
[b]*Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff, UK*
[c]*Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[d]*Experimental MRI Centre (EMRIC), Cardiff University School of Biosciences, Cardiff, UK*
[e]*Helen Durham Centre for Neuroinflammation, University Hospital of Wales, Cardiff, UK*
[f]*Department of Medical Physics, Leeds Teaching Hospitals NHS Trust, Leeds, UK*
[g]*Department of Medical Imaging and Medical Physics, Sheffield Teaching Hospital NHS Foundation Trust, Sheffield UK*
[h]*Institute for Advanced Biomedical Technologies (ITAB), University of Chieti-Pescara, Chieti, Italy*
[i]*Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Australia*

1. Estimation of SNR from in vivo data

To estimate the SNR of our in vivo diffusion sequence, we used the SNR_{mult} approach described by Dietrich et al. (2007). This relies on calculating the ratio between the mean of the signal divided by the standard deviation of the underlying (Gaussian) noise. To do this, the six non diffusion-weighted images were taken as repeated acquisitions. An SNR map was calculated for 29 MS patients, followed by registration of each map to MNI space. We averaged the SNR maps across participants. Thresholding the SNR maps at 20 showed that this value was exceeded across the whole brain, with the exception of the putamen. From this, we concluded that a SNR of 20 was a conservative estimate. Note that in the calculation, the noise was assumed Gaussian, which is the case for SNR > 3 (Gudbjartsson and Patz, 1995). However, the SNR was estimated based on non-diffusion weighted images, and the diffusion-weighted images have lower signal. To check that the signal attenuation was not more than 1/7 (which would cause the SNR in the attenuated images to fall below 20), we visualised in a few data sets the maximum signal attenuation, and in the vast majority of voxels this was not the case.

*Corresponding author: Ilona Lipp
Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany
Email address: lippi@cbs.mpg.de

[1]These authors have contributed equally to this manuscript.
[2]These authors have contributed equally to this manuscript.

Table 1: **Scan parameters**. All sequences were acquired at 3T. For each of the sequences, the main acquisition parameters are provided. **Acronyms:** FLAIR = fluid-attenuated inversion recovery, FSPGR = fast spoiled gradient echo, SE = spin-echo, IR = inversion recovery, EPI = echo-planar imaging, EFGRE = enhanced fast gradient echo, TE = echo time, TR = repetition time, TI = inversion time.

All at 3T	**T1-weighted**	**PD/T2-weighted**	**FLAIR (T2-weighted)**	**DTI**	**MT**
Pulse sequence(s)	FSPGR	SE	SE\IR	SE\EPI	EFGRE
Native resolution (mm^3)	1.0x1.0x1.0	0.94x0.94x 4.5	0.86x0.86x4.5	1.8x1.8x2.4	0.94x0.94x1.9
Field of view (mm)	256	240	220	230	240
Matrix size	256x256x172	256x256	256x256	96x96x36	128x128x100
Slices	none-3D	36 (3mm + 1.5mm gap)	36 (3mm + 1.5mm gap)	57	none-3D
Total acquisition time (min)	7.5	2	3	12.5	4.5
TE,TR (ms)	3.0,7.8	9.0/80.6,3000	122.3,9502	94.5,16000	1.8,26.7
TI (ms)	450	-	2250	-	-
off-resonance pulse					450 degrees, 2kHz off
Flip angle (degrees)	20	90	90	90	5

Table 2: **Comparison of FOD threshold for CSD in vivo.** As done before, for both CSD FOD thresholds (0.1 and 0.3) and each tissue type (Ctrl, NAWM, T2L, T1l), we calculated voxel-wise averages of the following parameters: the number of streamlines found in a voxel, the average FA / FOD amplitude across all streamlines found in a voxel, the number of streamlining processes stopped due to the amplitude threshold in a voxel, and the number of streamlining processes stopped due to the angle threshold in a voxel. The Mean ± SD of these measures across healthy controls (*Ctrl*; N = 19) and patients (*NAWM, T2L, T1L*; N = 29 of the MS subgroup) are reported. The values across different tissues types were statistically compared (unpaired t-test between Ctrl and NAWM tissue; paired t-tests for T2OL vs NAWM, and T1L vs T2OL).

	Ctrl	**NAWM**	**T2OL**	**T1L**	**NAWM vs Ctrl**	**T2OL vs NAWM**	**T1L vs T2OL**
Average number of streamlines per voxel							
CSD (0.1)	63.37 ± 10.22	66.12 ± 8.77	70.70 ± 13.18	70.69 ± 18.86	t = 0.99, p = .33	t = 2.4, p = .02	t = -0.01, p = .996
CSD (0.3)	63.36 ± 10.21	66.07 ± 8.75	70.68 ± 13.23	70.74 ± 18.72	t = 0.98, p = .33	t = 2.4, p = .02	t = -0.03, p = .98
Average peak amplitude (FOD amplitude / FA) of all streamlines per voxel							
CSD (0.1)	0.42 ± 0.04	0.43 ± 0.05	0.43 ± 0.06	0.42 ± 0.06	t = 0.78, p = .44	t = 0.64, p = .53	t = -2.2, p = .04
CSD (0.3)	0.42 ± 0.04	0.43 ± 0.05	0.43 ± 0.06	0.42 ± 0.06	t = 0.78, p = .44	t = 0.63, p = .53	t = -2.2, p = .04
Average number of stopped streamlines per voxel due to amplitude threshold							
CSD (0.1)	0.97 ± 0.03	0.87 ± 0.09	0.81 ± 0.19	0.80 ± 0.28	t = -4.7, p <.0001	t = -1.7, p = .10	t = -0.38, p = .71
CSD (0.3)	0.96 ± 0.03	0.87 ± 0.09	0.81 ± 0.19	0.80 ± 0.28	t = -4.7, p <.0001	t = -1.7, p = .10	t = -0.23, p = .82
Average number of stopped streamlines per voxel due to angle threshold							
CSD (0.1)	1.37 ± 0.09	1.41 ± 0.12	1.76 ± 0.59	2.04 ± 0.76	t = 1.1, p = .26	t = 3.2, p < .01	t = 2.9, p = .01
CSD (0.3)	1.37 ± 0.08	1.41 ± 0.12	1.76 ± 0.59	2.07 ± 0.80	t = 1.3, p = .27	t = 3.2, p < .01	t = 2.8, p = .01

Figure 1: **Example reconstruction of a CST and arcuate fasciculus.** The manual delineation of a left CST and a left arcuate fasciculus in a healthy control participant is shown. The placed gates (green: AND gate, blue: SEED gate) and the resulting fibre tracts are shown in all three planes. **Acronyms:** CST: cortico-spinal tract, Arc: arcuate fasciculus.

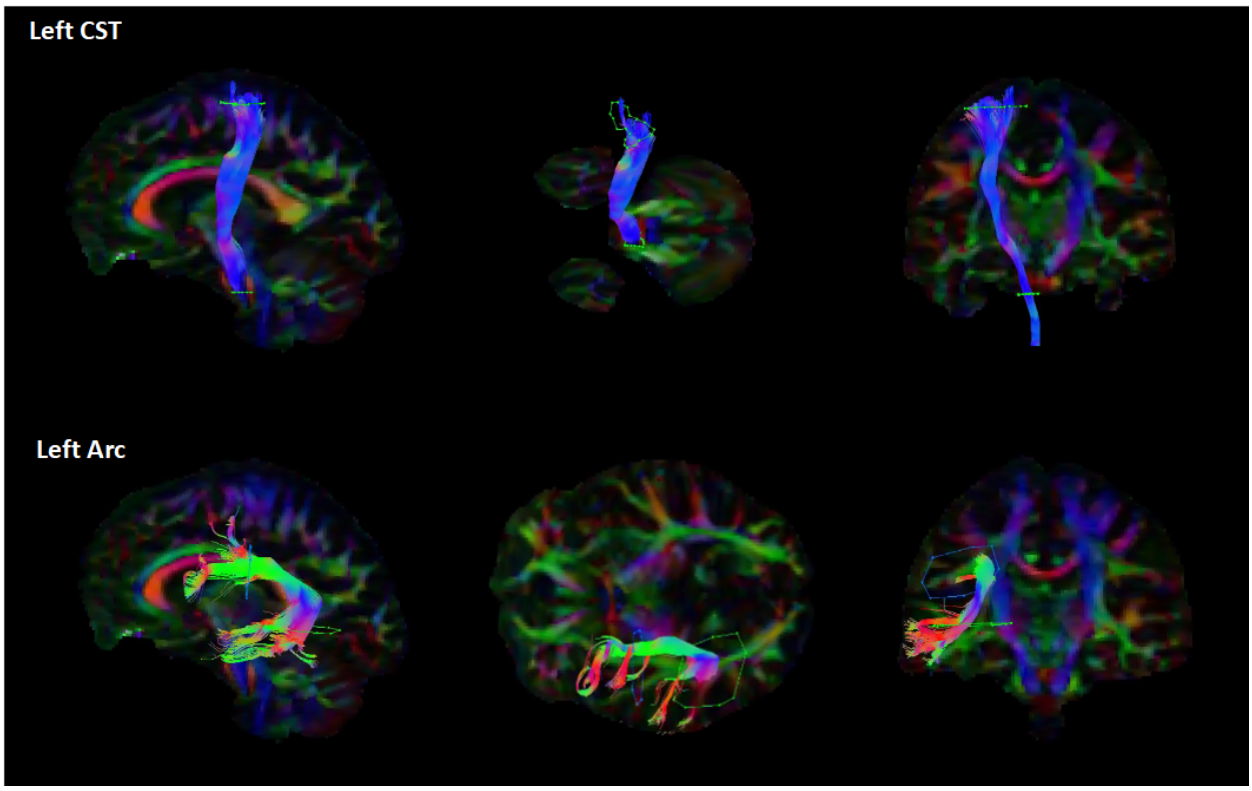

Figure 2: **Comparison of the three fibre orientation reconstruction algorithms in simulated data with b = 2000 s/mm^2.** Simulated substrates varied in their intracellular volume fraction (ICFV). The true fibre orientation of the parallel cylinders in each substrate was along the z-axis (single fibre population), or along the z- as well as the y-axis (crossing fibre populations). **A:** For each approach (tensor-based, dRL and CSD), we calculated the percentage of all voxels within each substrate type for which the 'true' underlying fibre configuration peak(s) could be detected. As a control, we also calculated this percentage for 'false' peaks (orthogonal to the true peak(s)). In each case, the left-most plot shows the FA for each substrate type. **B:** Dispersion across all detected peaks of a substrate type was calculated.

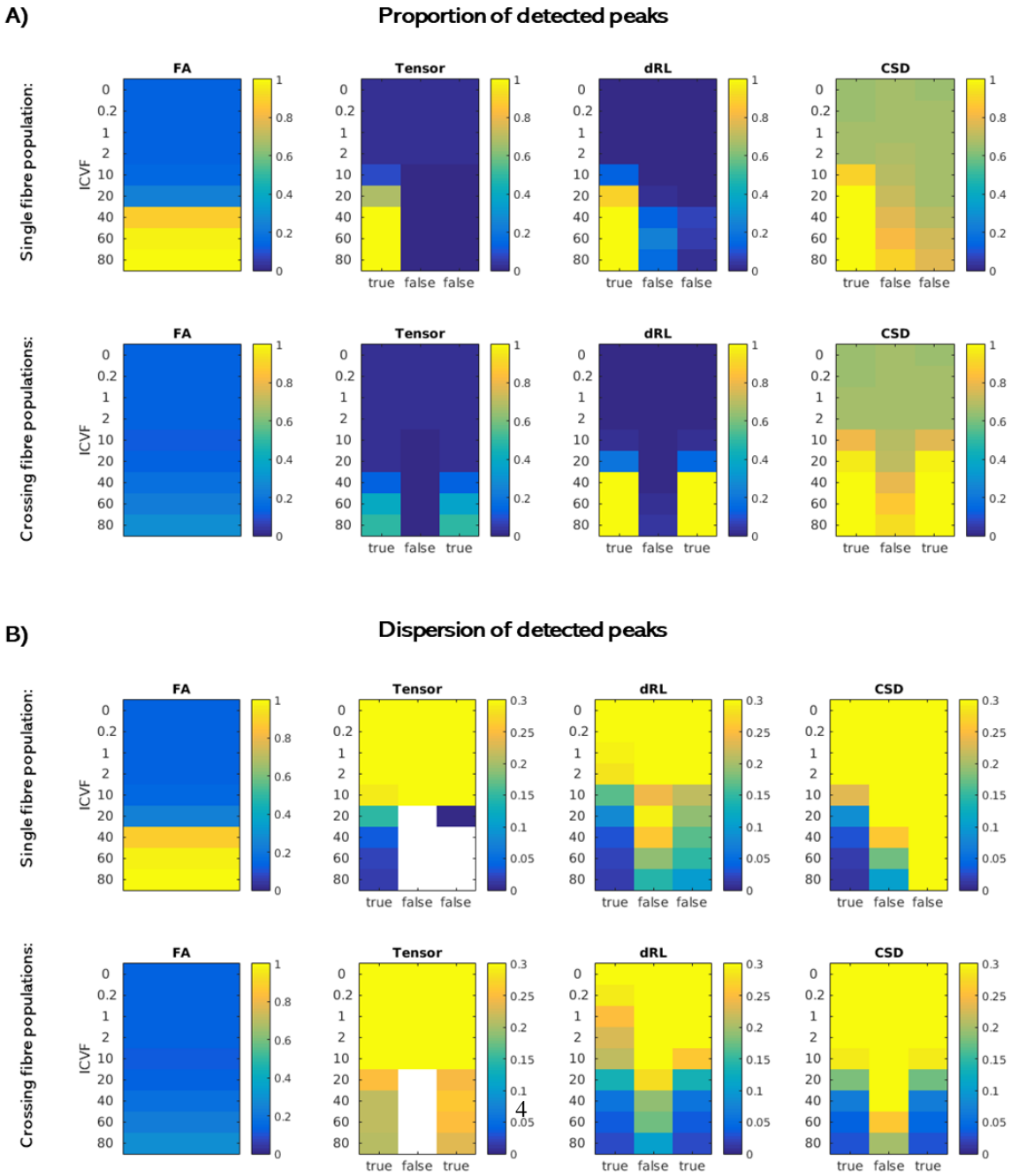

Figure 3: **Inter-operator anatomical agreement.** Spatial Dice coefficients were computed to quantify the overlap between segmented tracts from two independent operators in %. This was done for tracts from five healthy controls. Each time, boxplots are presented for each tract. **Acronyms:** l: left, r: right, CST: cortico-spinal tract, ARC: arcuate fasciculus.

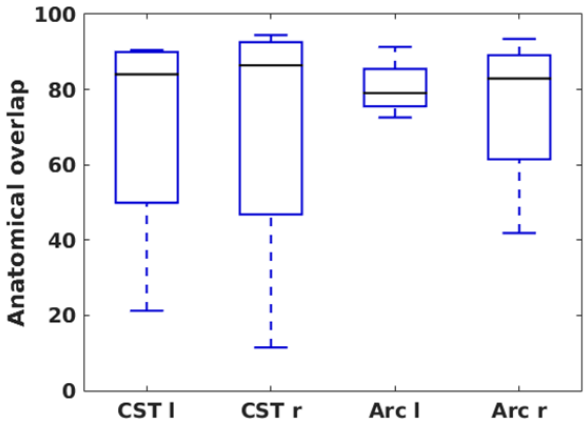

Figure 4: **Anatomical overlap between manually and automatically segmented tracts.** Spatial Dice coefficients were computed to quantify the overlap between manually and automatically segmented tracts in %. Each time, boxplots for the patient subgroup (red; N = 29 for CST r and Arc l, N = 28 for CST l, N = 26 for Arc r) and controls (blue; N = 19 for CST l, CST r, Arc l, N = 18 for Arc r) are presented. Outliers are indicated with black crosses **Acronyms:** l: left, r: right, CST: cortico-spinal tract, Arc: arcuate fasciculus.

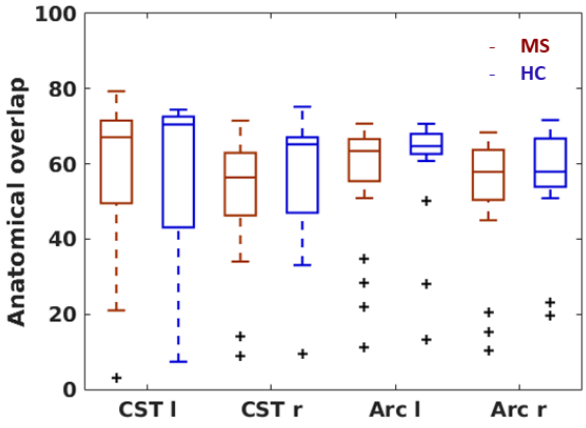

Table 3: **Correlations between tract-specific average microstructural metrics for automated vs manual tract dissection.** For each metric and tract and each group separately, the sample size (N) and the Pearson correlation coefficient and the corresponding p-values are reported. From Figure 5 it is evident that some low correlations may be caused by outliers. For these correlations, bivariate outliers were excluded (based on (Rousselet and Pernet, 2012)) and outlier-robust correlations and corresponding p-values are also reported. Note that correlations were not systematically lower in patients, even though the tracts from the healthy controls were used to create the model for automated tractography. **Acronyms:** l: left, r: right, CST: cortico-spinal tract, Arc: arcuate fasciculus, FA = fractional anisotropy, RD = radial diffusivity, MTR = magnetisation transfer ratio.

			FA			
	N MS	**N HC**	**MS**	**MS robust**	**HC**	**HC robust**
CST l	28	19	r =.63, $p < .001$	r =.96, $p < .001$	r =.86, $p < .001$	r =.86, $p < .001$
CST r	29	19	r =.51, $p = 0.01$	r =.85, $p < .001$	r =.91, $p < .001$	r =.91, $p < .001$
Arc l	29	19	r =.61, $p < .001$	r =.97, $p < .001$	r =.49, $p = 0.03$	r =.98, $p < .001$
Arc r	26	18	r =.56, $p < .001$	r =.94, $p < .001$	r =.68, $p < .001$	r =.88, $p < .001$
			RD			
	N MS	**N HC**	**MS**	**MS robust**	**HC**	**HC robust**
CST l	28	19	r =.51, $p = .01$	r =.95, $p < .001$	r =.91, $p < .001$	r =.93, $p < .001$
CST r	29	19	r =.52, $p < .001$	r =.89, $p < .001$	r =.86, $p < .001$	r =.91, $p < .001$
Arc l	29	19	r =.75, $p < .001$	r =.96, $p < .001$	r =.57. $p = 0.01$	r =.88, $p < .001$
Arc r	26	18	r =.76, $p < .001$	r =.98, $p < .001$	r =.72, $p < .001$	r =.96, $p < .001$
			MTR			
	N MS	**N HC**	**MS**	**MS robust**	**HC**	**HC robust**
CST l	28	19	r =.78, $p < .001$	r =.98, $p < .001$	r =.98, $p < .001$	r =.96, $p < .001$
CST r	29	19	r =.85, $p < .001$	r =.85, $p < .001$	r =.97, $p < .001$	r =.93, $p < .001$
Arc l	29	19	r =.66, $p < .001$	r =.99, $p < .001$	r =.92, $p < .001$	r =.85, $p < .001$
Arc r	26	18	r =.80, $p < .001$	r =.96, $p < .001$	r =.88, $p < .001$	r =.88, $p < .001$

Table 4: **Bland-Altman analysis**. Shown are the bias (mean of method difference), error (SD of method difference), lower and upper boundaries, and the Pearson correlation coefficient between mean and difference of the two methods. Data were collapsed across groups and hemispheres for this analysis. Data from 123 and 130 patients and 19 controls contributed to the data points for the left and right CST, and data from 130 and 122 patients and 19 controls contributed to the data points for the left and right Arc, respectively.

Tract	**Metric**	**Bias**	**Error**	**Lower limit**	**Upper limit**	***r(p)***	**N**
CST	FA	0.0857	0.0228	0.0411	0.1303	0.39 ($p < .0001$)	291
ARC	FA	0.0903	0.0190	0.0531	0.1276	0.40 ($p < .0001$)	290
CST	RD (in $10^{-3}m^2/s$)	-0.0788	0.0506	-0.1779	0.0203	-0.07 ($p = .25$)	291
ARC	RD (in $10^{-3}m^2/s$)	-0.0829	0.0243	-0.1305	-0.0352	0.25 ($p <.0001$)	290
CST	MTR	0.0236	0.0085	0.0069	0.0402	0.13 ($p = .02$)	290
ARC	MTR	0.0193	0.0054	0.0087	0.0299	0.35 ($p < .0001$)	290

Figure 5: **Correlations between tract-specific metrics for automated vs manual dissection.** For each metric (rows) and tract (columns), the scatterplot for the correlation between automated and manual approach is shown. Patients from the MS subgroup are represented by red dots, and healthy controls by blue dots. In each case, identified bivariate outliers are represented with the asterix symbol. **Acronyms:** l: left, r: right, CST: cortico-spinal tract, Arc: arcuate fasciculus, FA = fractional anisotropy, RD = radial diffusivity (in 10^{-3} m^2/s), MTR = magnetisation transfer ratio.

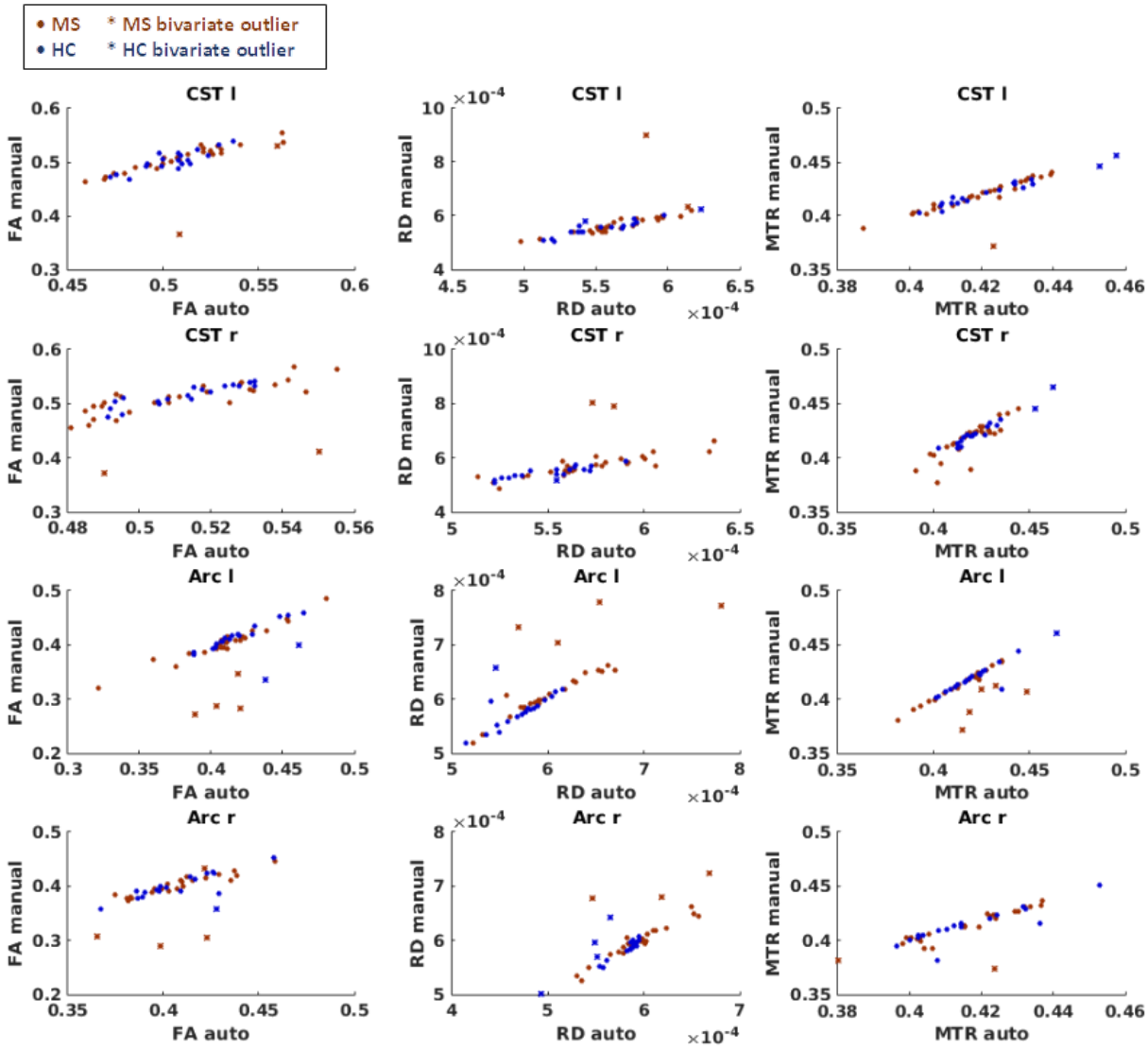

Figure 6: **Systematic differences between individually segmented and probability-based tracts.** For each patient and control, the difference in FA, RD, and MTR between individually and probability-based tracts were calculated. For each tract and metric, a boxplot is shown, also indicating the line of zero bias. The statistical group comparison of these difference measures is reported in Table **??** . Data from 123 and 130 patients and 19 controls contributed to the plot for the left and right CST, respectively, and data from 130 and 122 patients and 19 controls contributed to the plot for the left and right Arc, respectively. **Acronyms:** l: left, r: right, CST: cortico-spinal tract, Arc: arcuate fasciculus, FA = fractional anisotropy, RD = radial diffusivity (in 10^{-3} m^2/s), MTR = magnetisation transfer ratio.

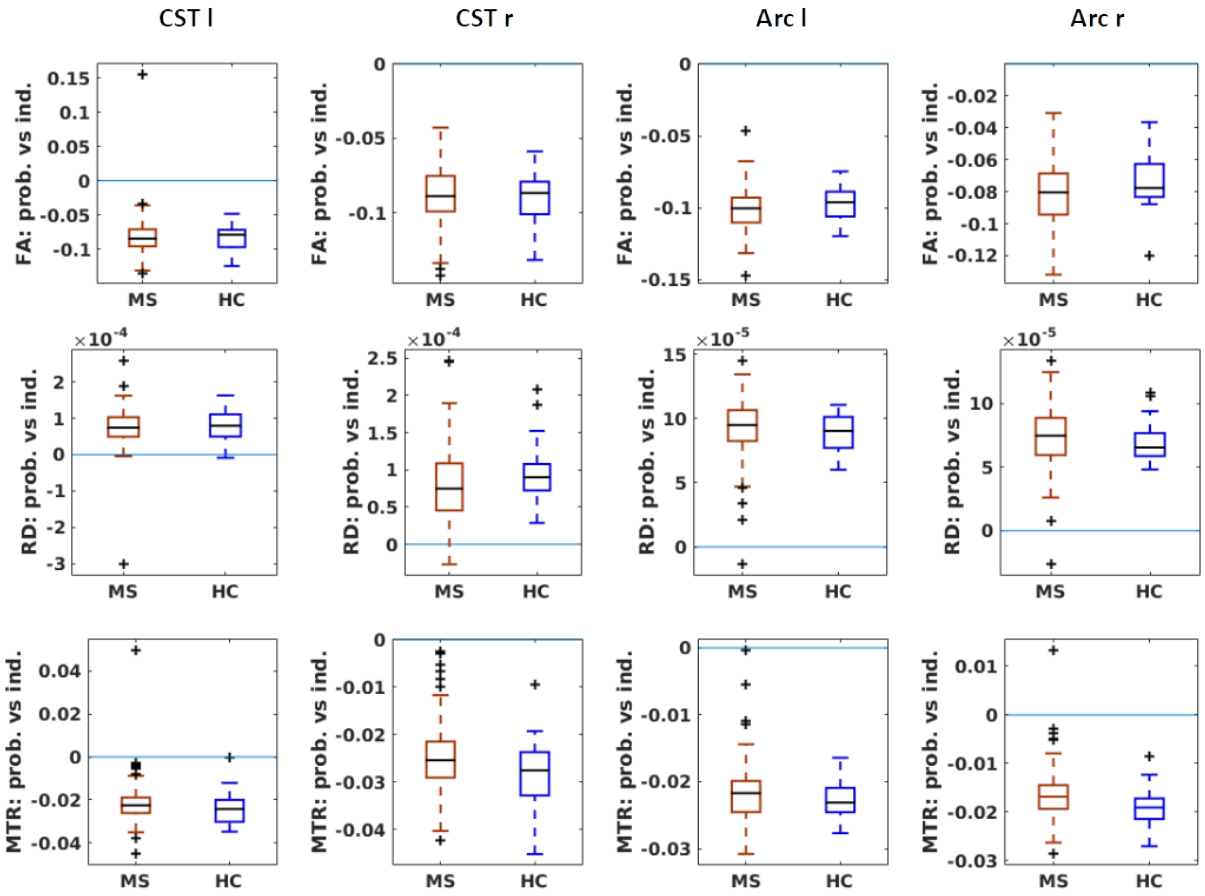

Figure 7: **Bland-Altman plots between tract-specific metrics for automated vs manual dissection.** Plotted are mean the metric across the two methods (probabilistic and individual) vs their difference. Data were collapsed across hemispheres for plotting. Pearson correlation coefficients are provided for each group and hemisphere separately. MS data points are plotted in red, HC in blue. Data from 123 and 130 patients and 19 controls contributed to the plot for the left and right CST, respectively, and data from 130 and 122 patients and 19 controls contributed to the plot for the left and right Arc, respectively. **Acronyms:** l: left, r: right, CST: cortico-spinal tract, Arc: arcuate fasciculus, FA = fractional anisotropy, RD = radial diffusivity (in 10^{-3} m^2/s), MTR = magnetisation transfer ratio.

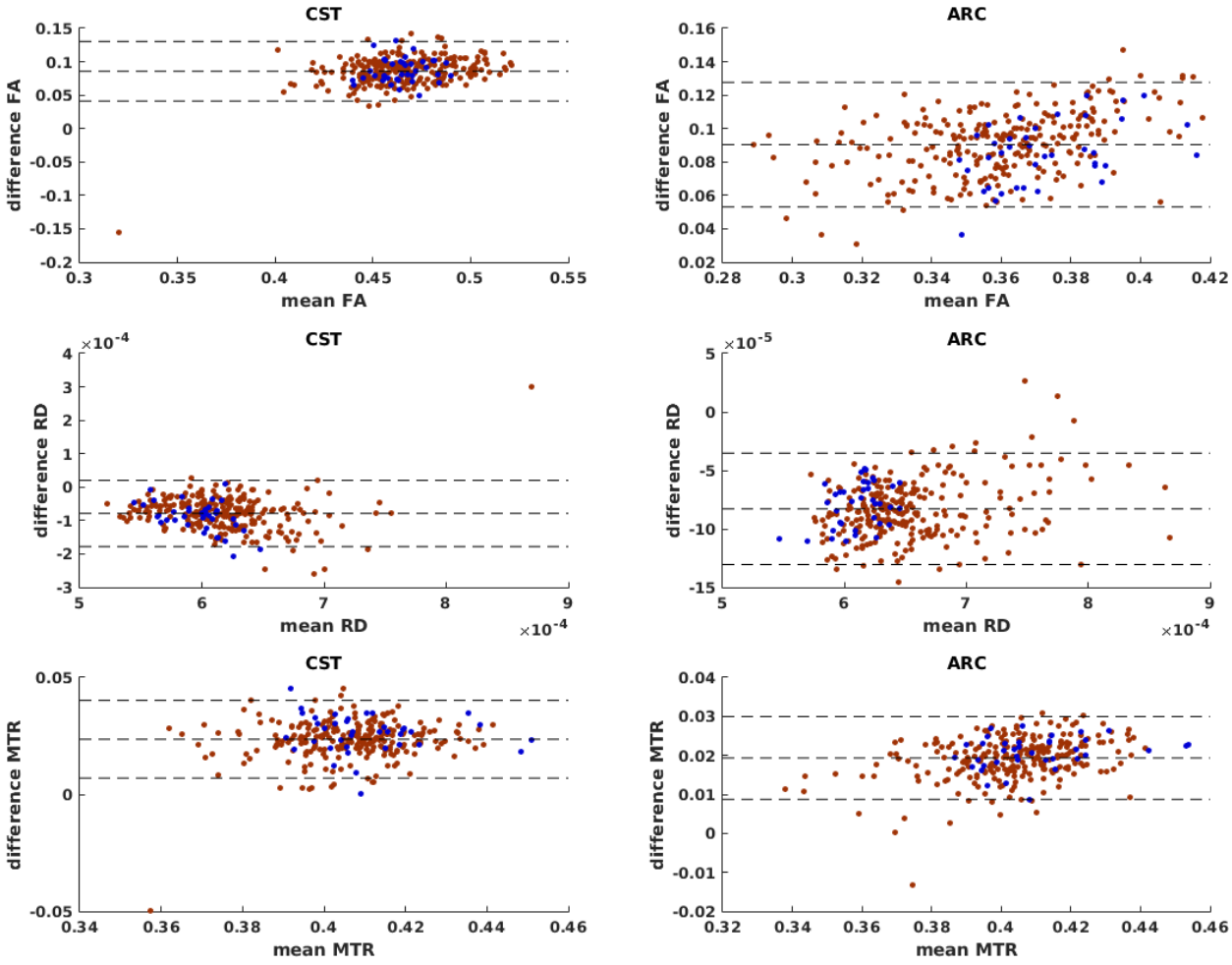

Table 5: **Demographic and clinical characteristics of patients whose CST / Arc could vs could not be reconstructed.** Characteristics are provided for all patients for whom both CSTs could be reconstructed (CST+) vs those for whom at least one reconstruction failed (CST-), for all patients for whom both Arcs could be reconstructed (Arc+) vs those for whom at least one reconstruction failed (Arc-). Patients in the CST- and Arc- groups did not overlap. Unless otherwise indicated, descriptive statistics provided are means and standard deviations. For statistical comparison between the groups (CST+ vs CST- and Arc+ vs Arc-), Chi-square tests were computed for categorical variables, Kruskall-Wallis tests for skewed variables (9 hole peg test and timed 25 foot walk), and unpaired *t*-tests for the rest. P values for group differences are provided. **Acronyms:** RR = Relapsing-remitting, P = progressive MS (includes primary and secondary progressive patients), EDSS = Extended Disability Status Scale, MSIS-29 = Multiple Sclerosis Impact Scale 29 items, 9-HPT: 9 hole peg test, T25-FW: timed 25 foot walk, PASAT = Paced Auditory Serial Addition Test (3 and 2 second version), DMT = disease-modifying treatment, BDI = Beck Depression Inventory, MFIS = Modified Fatigue Impact Scale. Normalized brain and grey matter volume was calculated using SIENAX (Smith et al., 2002).

Variable	**CST+**	**CST-**	**Arc+**	**Arc-**	**CST+ vs CST- (p)**	**Arc+ vs Arc- (p)**
N	123	8	121	10		
Age	44.4 ± 9.5	45.1 ± 7.8	44.6 ± 9.1	42.4 ± 12.0	.8343	.4731
Sex (F/M)	80/43	5/3	79/42	6/4	.8840	.7363
Education (years)	15.5 ± 3.9	18.1 ± 4.5	15.7 ± 3.9	14.3 ± 4.2	.0623	.2712
Disease duration (years)	12.2 ± 7.6	16.1 ± 4.9	12.3 ± 7.3	14.2 ± 9.9	.1489	.4340
Disease course (RR/P)	100/23	5/3	95/26	10/0	.1964	.1016
EDSS (median/iqr)	4.0 ± 1.4	4.0 ± 3.2	4.0 ± 1.5	4.0 ± 0.2	.6372	.9460
MSIS-29	65.2 ± 29.7	73.0 ± 27.5	66.6 ± 29.9	54.3 ± 21.9	.4673	.2054
9-HPT (right) in sec. (across 2 trials) median/iqr	22.3 ± 6.5	27.9 ± 8.5	22.5 ± 6.9	22.9 ± 7.4	.1194	.7815
T25-FW in sec. (across 2 trials) median/iqr	8.1 ± 8.9	13.9 ± 18.8	8.6 ± 10.1	6.1 ± 2.6	.1062	.4473
PASAT 3s	40.0 ± 13.9	38.2 ± 17.1	39.8 ± 13.9	41.2 ± 16.7	.7340	.7607
PASAT 2s	27.7 ± 11.7	23.9 ± 10.8	27.4 ± 11.2	28.2 ± 16.7	.3679	.8378
DMT (Yes/No)	0.3 ± 0.5	0.1 ± 0.4	0.3 ± 0.5	0.4 ± 0.5	.2094	.6183
Depression (BDI)	12.1 ± 10.1	18.6 ± 14.5	12.7 ± 10.4	10.2 ± 10.6	.0864	.4626
Fatigue (MFIS)	39.0 ± 20.8	49.9 ± 15.6	40.3 ± 20.6	31.5 ± 21.1	.1491	.1958
Normalized GM volume (cm^3)	596.3 ± 64.3	581.4 ± 43.6	596.8 ± 61.7	579.0 ± 81.1	.5183	.3950
Normalized whole brain volume (cm^3)	1175.5 ± 117.8	1143.4 ± 82.0	1175.2 ± 116.6	1153.5 ± 111.2	.4504	.5717
T2-hyperintense lesion volume (cm^3)	4.3 ± 4.7	2.8 ± 3.2	4.0 ± 4.5	5.8 ± 5.8	.3992	.2620

Table 6: **Group differences between microstructural measures in probability-based and individually segmented tracts.** For each patient and control and for each tract, Mean ± SD microstructural metrics (FA, RD and MTR) were extracted from the probability-based as well as for the individually dissected tract mask. For each tract and metric, an unpaired t-test between the values from healthy controls (HC) and patients (MS) was calculated, and t and p statistics are provided for each comparison. Data from 123 and 130 patients and 19 controls contributed to the analysis for the left and right CST, and data from 130 and 122 patients and 19 controls contributed to the analysis for the left and right Arc, respectively. **Acronyms:** l: left, r: right, CST: cortico-spinal tract, Arc: arcuate fasciculus, FA = fractional anisotropy, RD = radial diffusivity (in 10^{-3} m^2/s), MTR = magnetisation transfer ratio.

	FA					
	Probability-based			Individual		
Tract	HC (mean±SD)	MS (Mean±SD)	HC vs MS	HC (Mean±SD)	MS (Mean±SD)	HC vs MS
CST l	0.42±0.02	0.42±0.02	t=-0.37, p=0.7112	0.50±0.02	0.51±0.04	t=-0.26, p=0.7923
CST r	0.42±0.01	0.42±0.02	t=-0.07, p=0.9477	0.51±0.01	0.51±0.03	t=0.39, p=0.6937
ARC l	0.32±0.02	0.31±0.02	t=3.13, p=0.0021	0.42±0.02	0.41±0.03	t=1.68, p=0.0945
ARC r	0.34±0.01	0.32±0.02	t=3.16, p=0.0019	0.41±0.02	0.40±0.03	t=1.25, p=0.2124
	RD (in 10^{-3} m^2/s)					
	Probability-based			Individual		
Tract	HC (Mean±SD)	MS (Mean±SD)	HC vs MS	HC (Mean±SD)	MS (Mean±SD)	HC vs MS
CST l	0.64±0.04	0.65±0.05	t=-1.48, p=0.1419	0.56±0.03	0.58±0.06	t=-1.81, p=0.0722
CST r	0.65±0.04	0.66±0.05	t=-0.91, p=0.3655	0.55±0.02	0.58±0.04	t=-3.01, p=0.0031
ARC l	0.66±0.02	0.71±0.05	t=-3.94, p=0.0001	0.57±0.03	0.61±0.06	t=-3.15, p=0.0020
ARC r	0.64±0.02	0.68±0.05	t=-3.80, p=0.0002	0.57±0.03	0.61±0.06	t=-2.89, p=0.0045
	MTR					
	Probability-based			Individual		
Tract	HC (Mean±SD)	MS (Mean±SD)	HC vs MS	HC (Mean±SD)	MS (Mean±SD)	HC vs MS
CST l	0.40±0.01	0.40±0.01	t=1.12, p=0.2643	0.42±0.01	0.42±0.02	t=1.49, p=0.1374
CST r	0.40±0.02	0.39±0.01	t=0.64, p=0.5204	0.42±0.01	0.42±0.01	t=1.64, p=0.1039
ARC r	0.40±0.02	0.39±0.02	t=1.37, p=0.1722	0.42±0.02	0.42±0.02	t=1.48, p=0.1399
ARC r	0.40±0.02	0.39±0.02	t=1.49, p=0.1379	0.42±0.02	0.41±0.02	t=1.86, p=0.0648

References

Dietrich, O., Raya, J.G., Reeder, S.B., Reiser, M.F., Schoenberg, S.O., 2007. Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. Journal of Magnetic Resonance Imaging 26, 375–385. doi:10.1002/jmri.20969.

Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI Data. MRM 34, 910–914.

Rousselet, G.A., Pernet, C.R., 2012. Improving standards in brain-behavior correlation analyses. Frontiers in Human Neuroscience 6, Article 119. doi:10.3389/fnhum.2012.00119.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., Stefano, N.D., 2002. Accurate, robust, and automated longitudinal and cross-cectional brain change analysis. NeuroImage 489, 479–489. doi:10.1006/nimg.2002.1040.