

MICRA: Microstructural image compilation with repeated acquisitions

Kristin Koller^{a,*}, Umesh Rudrapatna^a, Maxime Chamberland^a, Erika P. Raven^a,
Greg D. Parker^{a,c}, Chantal M.W. Tax^a, Mark Drakesmith^a, Fabrizio Fasano^d, David Owen^a,
Garin Hughes^a, Cyril Charron^a, C John Evans^a, Derek K. Jones^{a,b}

^a Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, CF24 4HQ, United Kingdom

^b Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Australia

^c Experimental MRI Centre (EMRIC), School of Biosciences, Cardiff University, Cardiff, United Kingdom

^d Siemens Healthcare GmbH, Erlangen, Germany

A B S T R A C T

We provide a rich multi-contrast microstructural MRI dataset acquired on an ultra-strong gradient 3T Connectom MRI scanner comprising 5 repeated sets of MRI microstructural contrasts in 6 healthy human participants. The availability of data sets that support comprehensive simultaneous assessment of test-retest reliability of multiple microstructural contrasts (i.e., those derived from advanced diffusion, multi-component relaxometry and quantitative magnetisation transfer MRI) in the same population is extremely limited. This unique dataset is offered to the imaging community as a test-bed resource for conducting specialised analyses that may assist and inform their current and future research. The Microstructural Image Compilation with Repeated Acquisitions (MICRA) dataset includes raw data and computed microstructure maps derived from multi-shell and multi-direction encoded diffusion, multi-component relaxometry and quantitative magnetisation transfer acquisition protocols. Our data demonstrate high reproducibility of several microstructural MRI measures across scan sessions as shown by intra-class correlation coefficients and coefficients of variation. To illustrate a potential use of the MICRA dataset, we computed sample sizes required to provide sufficient statistical power *a priori* across different white matter pathways and microstructure measures for different statistical comparisons. We also demonstrate whole brain white matter voxel-wise repeatability in several microstructural maps. The MICRA dataset will be of benefit to researchers wishing to conduct similar reliability tests, power estimations or to evaluate the robustness of their own analysis pipelines.

1. Introduction

The primary aim of this work was to collect, and disseminate to the neuroimaging community, the MICRA (Microstructural Image Compilation with Repeated Acquisitions) data set - a unique and rich multi-variate (diffusion, relaxometry, magnetisation transfer) microstructural MRI archive that allows variance and co-variance of measures to be estimated between tracts, between multiple time-points and between different individuals.

To provide just one example of the utility of such a dataset, we present estimates of sample size calculations that could inform current or planned future microstructural imaging experiments. With the movement towards “open science” practices (Allen and Mehler, 2019; Munafò et al., 2017), there is increasing demand to demonstrate *a priori* that study designs are adequately powered to answer a targeted question. In turn, this requires an assessment of test-retest repeatability as input to the sample size estimations. There is also a trend to complement diffusion-based microstructural measurements with additional measures that have enhanced sensitivity to myelin, including those derived from relaxometry and magnetisation transfer-based approaches (Ercan et al.,

2018; Friedrich et al., 2020; Geeraert et al., 2019; Jung et al., 2018; Lipp et al., 2019; Metzler-Baddeley et al., 2017; Morris et al., 2020; Uddin et al., 2019). While it is appealing to collect data across the gamut of *available* options, each new contrast takes time (and invokes real cost) to acquire, and it is important to establish that the measurements are reproducible in themselves, and to establish their reproducibility so that, if necessary, one might prioritise certain image contrasts over others. There is quite an extensive literature on the reproducibility of single neuroimaging measures (e.g., Papinutto et al., 2013; Stikov et al., 2015; Biswal et al., 2010; O'Connor et al., 2017; Vollmar et al., 2010; Grech-Sollars et al., 2015; Prčkovska et al., 2016; Tong et al., 2019), focusing predominantly on inter-site reproducibility and with some focusing on within-site reproducibility. However, to the best of our knowledge, comparison of the reproducibility of multiple microstructural imaging measures, measured in the same population, does not exist at the time scale and sampling frequency presented here.

The resource provided here to the community addresses this gap, allowing for detailed assessment of the reliability of measures derived from optimised multi-shell diffusion, multi-component relaxometry and quantitative magnetisation transfer acquisition protocols. Here we com-

* Corresponding author. Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, CF24 4HQ, UK.
E-mail address: koller@cardiff.ac.uk (K. Koller).

pute reliability statistics (intra-class correlation and coefficient of variation) across three example tracts and individual measures from each microstructural imaging approach. Although these are illustrative examples, these data could be reprocessed and used to compute other parameters from diffusion, relaxometry and quantitative magnetisation transfer (QMT) models. We provide protocols that can be used for power calculations highlighting the utility of the resource to researchers wishing to conduct similar reliability tests/ power calculations. However, this rich, high quality data resource, acquired on an ultra-strong-gradient Connectom 3T system that may not otherwise be readily accessible, will also be of value to those developing and evaluating new data-processing approaches (e.g., denoising, clustering, segmentation, joint-estimation and tractography algorithms).

2. Method

2.1. Participants

Six neurologically healthy adults (age range 24–30, 3 males and 3 females) were recruited from Cardiff University's staff and student panels. Screening for safety eligibility to undergo MRI scanning was conducted and participants received monetary compensation for participation. All participation was contingent upon prior written informed consent and ethical approval for this study was granted by Cardiff University's School of Psychology ethics committee.

2.2. MRI hardware: ultra-strong gradient 3T

Whole brain MRI data were acquired using an ultra-strong gradient (300mT/m) 3T Connectom research only MRI scanner, a modified 3T MAGNETOM Skyra (Siemens Healthcare, Erlangen, Germany). Compared to standard MRI gradients (45–80mT/m), the Connectom gradients allow for shorter diffusion times for a given diffusion weighting resulting in shorter minimum TEs (greater signal to noise ratio) and increased sensitivity to small water displacements (Jones et al., 2018; Setsompop et al., 2013).

2.3. MRI data acquisition

Each MRI session lasted approximately 45 min (CHARMED = 18 min, QMT = 12 min, McDESPOT = 11 min) and was repeated 5 times within a two-week period. Care was taken to avoid potential diurnal effects by performing scans for each participant at approximately the same time of day (i.e., within 1–2 h of the same scan start time-of-day).

The MRI protocol included the following sequences:

- (i) Multi-shell diffusion-weighted MRI: single-shot spin echo, echo planar imaging data were acquired with both anterior-posterior (AP) and posterior-anterior (PA) phase-encoded directions. The AP-encoded data comprised of two shells of 20 directions (uniformly-distributed according to Jones et al., 1999) at $b = 200$ s/mm² and 500 s/mm², one shell of 30 directions at $b = 1200$ s/mm² and three shells of 61 directions at each of $b = 2400$ s/mm², 4000 s/mm² and 6000 s/mm², with two leading non-diffusion-weighted images and a further 11 non-diffusion-weighted images, starting at the 33rd volume, and repeating every 20th volume thereafter. In the PA-encoded data, two non-diffusion-weighted images were acquired. The field of view was 220×220 mm in plane, the matrix size was $110 \times 110 \times 66$, reconstructed to a $110 \times 110 \times 66$ image resulting in $2 \times 2 \times 2$ mm³ isotropic voxels. The TR and TE were 3000 ms and 59 ms, respectively (for all b-values), and the diffusion gradient duration and separation were 7 ms and 24 ms, respectively.
- (ii) Multi-component relaxometry: data were acquired thanks to prototype sequences implementing the McDESPOT protocol (Deoni et al., 2008) with the following parameters: FOV: $220 \times 220 \times 178.88$,

matrix size: $128 \times 128 \times 104$ and $1.72 \times 1.72 \times 1.72$ mm³ isotropic voxels, SPGR: TR: 4 ms, TE: 1.9 ms, 8 flip angles (3, 4, 5, 6, 7, 9, 13 and 18°), SPGR-IR: TR: 4 ms, TE: 1.9 ms, TI: 450 ms, Flip angle: 5°, full k-space acquisition in PE and slice directions, SSFP: TR: 4.6 ms, TE: 2.3 ms, Flip angles (10, 13.33, 16.67, 20, 23.33, 30, 43.33 and 60°), SSFP180: as SSFP, but with 180° RF phase increments every TR.

- (iii) Optimised Quantitative Magnetisation Transfer (QMT, Mougin et al., 2010) data were acquired by using a prototype turbo-flash sequence with parameters: FOV: $220 \times 220 \times 178.88$, matrix size: $128 \times 128 \times 104$, resolution: $1.72 \times 1.72 \times 1.72$ mm³ isotropic voxels isotropic, turbo factor 4, radial reordering, non-selective excitation MT pulse duration: 15.36 ms, 11 MT-weighted volumes and 1 vol. without MT-weighting, 11 Frequency offsets (Hz) and 11 flip angles (degrees): 47,180 (628); 56,360 (332); 12,060, (628); 1000 (332); 1000 (333); 2750 (628); 2770 (628); 2790 (628); 2890 (628); 1000 (628); 1000 (628) (Cercignani and Alexander, 2006).

2.4. MRI data pre-processing and processing

Multi-shell diffusion-weighted data were pre-processed using a custom in-house pipeline comprising tools from both the FSL (Andersson et al., 2003; Andersson and Sotiropoulos, 2016) and MRTrx (Veraart et al., 2016; Vos et al., 2017) software packages and in-house software. Specifically, AP- and PA-encoded images were separately denoised (MRTrx dwidenoise, Veraart et al., 2016) and drift corrected (Vos et al., 2017), then merged (with incorporated EPI, susceptibility and motion correction; FSL *topup* (Andersson et al., 2003) and eddy (Andersson et al., 2016)) corrected for gradient non-linearity distortions (Glasser et al., 2013) with spatio-temporal b-value/vector tracking (Rudrapatna et al., 2018), and finally corrected for Gibbs ringing artefacts (MRTrx mrdegibbs, Kellner et al., 2016). Subsequent processing involved computation of: (i) free-water corrected fractional anisotropy (FA), mean diffusivity (MD) and radial diffusivity (RD) maps (Hoy et al., 2014) from diffusion tensor MRI using the $b = 1000$ s/mm² shell (linear least squares estimation with outlier rejection, Chang et al., 2005). Fibre Orientation Distribution Functions (fODFs) were derived from multi-shell multi-tissue Constrained Spherical Deconvolution (MSMT-CSD, Jeurissen et al., 2014). Microstructural parameters were estimated from all diffusion shells using the CHARMED model (Assaf et al., 2004) using a nonlinear regression routine employing the Levenberg–Marquardt optimization algorithm.

For McDESPOT data, motion correction was applied to the SPGR and SSFP data using FSL mcFLIRT (Jenkinson et al., 2002) followed by brain extraction (Smith, 2002). The QUIT toolbox (Wood, 2018) was utilised for all subsequent fitting. The DESPOT2-FM model was fitted to estimate a B0 map (Deoni, 2009), which was used as input for a final fitting to the 3-pool mcDESPOT model (Deoni et al., 2013), modelling myelin, extra-cellular and CSF contributions using the 'qimcdespot' function in QUIT.

QMT data were processed using the QUIT (Wood, 2018) toolbox using the Ramani model (Ramani et al., 2002). For QMT, the MT-weighted volumes were aligned to the non-MT contrast for motion correction and bias correction with B1 maps were applied by computing the B1 field correction based on the field estimate from the fifth MT volume, which was subsequently applied to all MT volumes (FSL FAST, Zhang et al., 2001).

2.5. White matter microstructure measures

The following microstructural measures were computed in each voxel: restricted diffusion signal fraction (RSF) fitted from CHARMED (nonlinear regression routine employing the Levenberg–Marquardt opti-

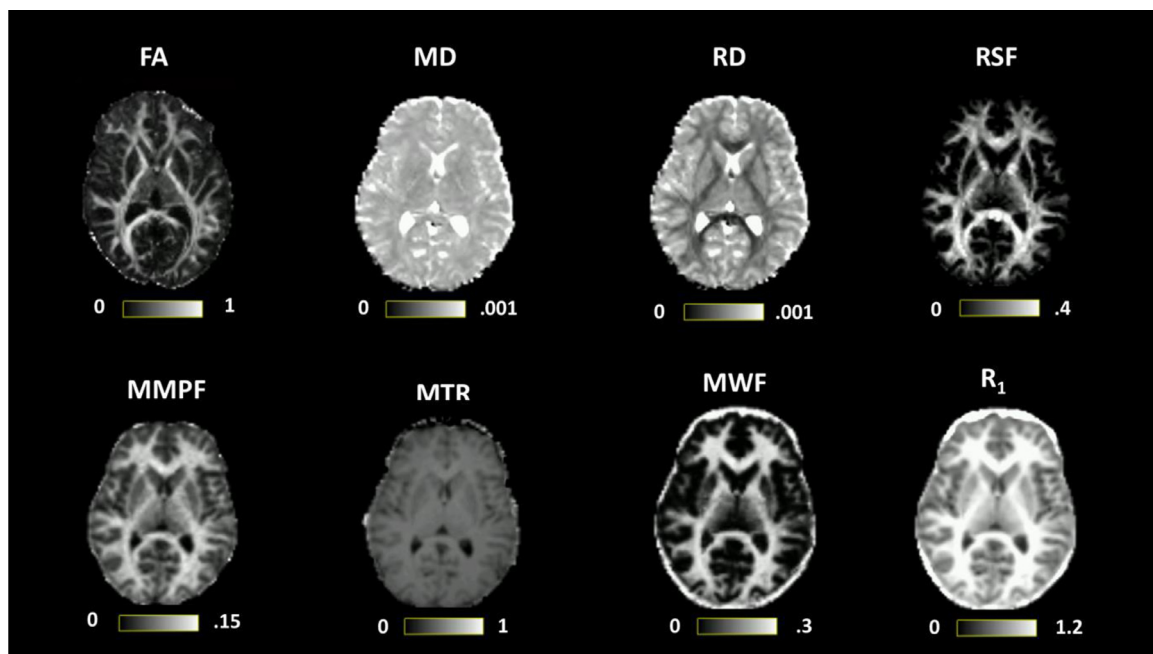


Fig. 1. Microstructural maps including FA (fractional anisotropy), MD (mean diffusivity, units = mm^2/s), RD (radial diffusivity, units = mm^2/s), RSF (restricted diffusion signal fraction), R_1 (longitudinal relaxation rate, units = s^{-1}), MWF (myelin water fraction), MMPF (macromolecular proton fraction) and MTR (magnetisation transfer ratio) computed for one representative participant. All contrasts are registered to diffusion space.

mization algorithm, Assaf et al., 2004), fractional anisotropy (FA), mean diffusivity (MD) and radial diffusivity (RD) from diffusion tensor MRI; myelin water fraction (MWF) and longitudinal relaxation rate (R_1) from the McDESPOT pipeline (Deoni and Kolind, 2015), the macromolecular proton fraction (MMPF) fitted from the QMT pipeline (Wood, 2018; Ramani et al., 2002) and magnetisation transfer ratio (MTR) computed using home-grown code. Quantitative maps were subsequently linearly registered to the space in which the diffusion MRI data were acquired ('native space') using FSL FLIRT (see Fig. 1 for illustration of all maps). Selections of models here are illustrative. The magnitude-reconstructed raw data are also included in the MICRA dataset, enabling researchers to explore other modelling options.

2.6. Virtual dissection of tracts

To assess test-retest repeatability, a white matter projection tract (cortico-spinal), association tract (arcuate fasciculus) and the fornix were virtually dissected from whole brain white matter maps for each participant at each time point with probabilistic tractography (MR-Trix iFOD2, 1000 seeds x 5000 streamlines, step size = $0.5 \times 2 \text{ mm}^3$ voxel size, angular threshold = $90^\circ \times$ step size/voxel size, FODF threshold = 0.05, Jeurissen et al., 2014, Fig. 2). The fornix was virtually dissected by placing region of interest masks in the anterior hippocampus and fornix body. The CST and the arcuate fasciculus were dissected using TractSeg (Wasserthal et al., 2018) using code available at <https://github.com/Wasserthal/TractSeg>.

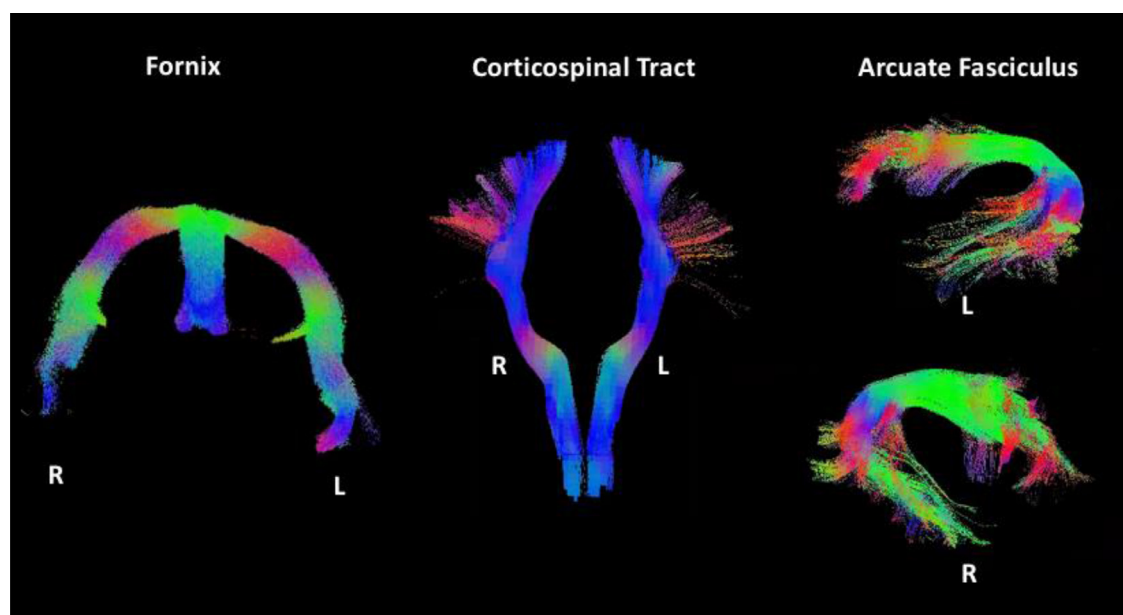


Fig. 2. Virtual dissections of the fornix, cortico-spinal tract and arcuate fasciculus dissections in one representative participant (see Methods).

Table 1

Test-retest reliability statistics for microstructure measures. Units (mean and SD) for MD and RD = 10^{-3} mm²/s; units for R_1 = s⁻¹, CV = averaged within subject, ICC = averaged across 5 repeated measures.

Metric	Tract	Mean	SD	CV (%)	ICC	ICC pval	r
FA	Fornix	0.47	0.005	1.0	0.93	<0.0001	0.82
MD		1.00	0.010	0.9	0.94	<0.0001	0.87
RD		1.00	0.004	0.6	0.96	<0.0001	0.88
RSF		0.13	0.003	2.1	0.88	<0.0001	0.65
MMPF		0.05	0.001	2.6	0.95	<0.0001	0.85
MWF		0.07	0.003	4.2	0.95	<0.0001	0.87
R_1		0.77	0.005	0.7	0.97	<0.0001	0.90
MTR		0.35	0.0004	0.1	0.94	<0.0001	0.77
FA	CST	0.55	0.003	0.6	0.97	<0.0001	0.89
MD		0.41	0.002	0.5	0.95	<0.0001	0.81
RD		0.63	0.001	0.2	0.95	<0.0001	0.78
RSF		0.27	0.003	1.2	0.86	0.002	0.52
MMPF		0.09	0.001	1.5	0.97	0.0003	0.74
MWF		0.16	0.003	1.7	0.97	<0.0001	0.95
R_1		1.04	0.004	0.4	0.98	<0.0001	0.91
MTR		0.41	0.002	0.5	0.78	0.01	0.37
FA	Arcuate Fasciculus	0.47	0.005	1.0	0.95	<0.0001	0.86
MD		0.47	0.005	1.0	0.96	<0.0001	0.85
RD		0.64	0.003	0.5	0.96	<0.0001	0.81
RSF		0.26	0.004	1.7	0.82	0.005	0.72
MMPF		0.10	0.002	2.1	0.92	<0.0001	0.79
MWF		0.16	0.002	1.4	0.98	<0.0001	0.93
R_1		1.07	0.001	0.1	0.97	<0.0001	0.90
MTR		0.42	0.003	0.6	0.97	<0.0001	0.92

One-way ANOVAs demonstrated no significant differences between ICC computed for individual time point pairs or between average ICC and individual time point ICC comparisons for the CST ($F(9,70) = 0.74$, $p = .67$) and the Fornix ($F(9,70) = 0.19$, $p = .07$). Although initially a difference was suggested for the Arcuate Fasciculus ($F(9,70) = 2.2$, $p = .03$), no comparisons survived correction for multiple comparisons as shown with Tukey post-hoc correction.

[//github.com/MIC-DKFZ/TractSeg/](https://github.com/MIC-DKFZ/TractSeg/). Tractography was conducted in each individual subject's diffusion space. Track density maps (TDI, Calamante et al., 2010) of the resultant tracts were computed and thresholded to exclude voxels through which streamlines passed less than 20 percent. As an *a priori* choice, the analysis was restricted to three tracts in order to show a demonstration of repeatability in one association, one projection and one commissural pathway.

2.7. Repeatability at the tract-level

Measures were extracted for each vertex in each streamline and averaged along each tract for statistical comparison. The intra-class correlation coefficient (two-way mixed, absolute agreement) and coefficient of variation were computed for average assessment of test-retest repeatability (Table 1) across all repeated scans.

Moreover, to ascertain whether there was an effect of time on reproducibility (i.e., do those measurements that are more closely-spaced in time agree better than those spaced further apart in time?), intra-class correlation coefficients were also computed for individual time point pairs across all scan sessions.

2.8. Repeatability at the voxel level

While our strong preference for microstructural comparisons is to use a 'tractometry'-based approach, (Bells et al., 2011; Chamberland et al., 2019) to perform individual/ group microstructural comparisons, we recognise the prevalence of voxel-based analyses. We therefore conducted a separate analysis of the reproducibility of each metric at the voxel-level within white matter across the whole brain. This was done by adopting the white matter skeletonisation approach popularised in the TBSS (Tract-Based Spatial Statistics, Smith et al., 2006, framework, part of FSL (Smith et al., 2004). First, and as above, the MMPF, R_1 , MTR and

MWF maps for a given participant and timepoint were first registered to the individual's native diffusion space using FLIRT (Jenkinson and Smith, 2001).

Then, FA maps from all subjects at all time-points were aligned into a common space using the nonlinear registration tool FNIRT (Andersson et al., 2007; Andersson et al., 2007) which uses a b-spline representation of the registration warp field (Rueckert, 1999). Next, the mean FA image was created and thinned to create a mean FA skeleton which represents the centres of all tracts common to the group. Each subject's aligned FA map was then projected onto this skeleton in MNI space. The nonlinear warps and skeleton projections generated for FA were applied to the corresponding non-FA maps (already in diffusion space) to create white matter skeletons in MNI space for these additional metrics. Prior to analysis, a further thresholding step was applied. Specifically, each voxel in the skeletonised data was only retained for further analysis if, in that voxel, all 5 participants at all six time-points had an FA > 0.2. This was to provide enhanced assurance that the analysis was restricted to white matter. For each metric, the Pearson correlation was then conducted across all voxels in the thresholded skeleton between each possible pair of time-points to assess the repeatability across whole brain white matter.

2.9. Demonstration of sample size estimation

To further illustrate the utility of the reproducibility data, we consider the minimum number of subjects needed to reach statistical power of 0.8 and significance of $\alpha = 0.05$ for two different types of statistical tests routinely carried out in the neuroimaging literature, as outlined below (power calculations were computed using G*Power, see Supplementary 2, Faul et al., 2009):

- Independent groups t-test* (e.g., for comparing a group of patients to a group of healthy controls). Here, we evaluate the numbers

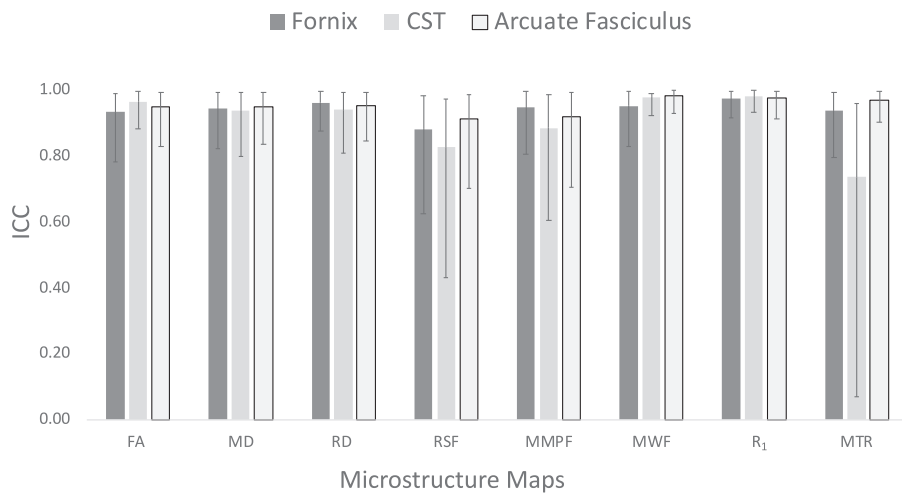


Fig. 3. Intra-class correlation coefficients (two-way mixed, absolute agreement) for test-retest repeatability of microstructure measures measured 5 times in 6 participants. ICC = intra-class correlation, CST = corticospinal tract, FA = fractional anisotropy, MD = mean diffusivity, RD = radial diffusivity, RSF = restricted diffusion signal fraction, MMPF = macromolecular proton fraction, MWF = myelin water fraction, R₁ = longitudinal relaxation rate, MTR = magnetisation transfer ratio.

Figure 4.1. Fractional Anisotropy

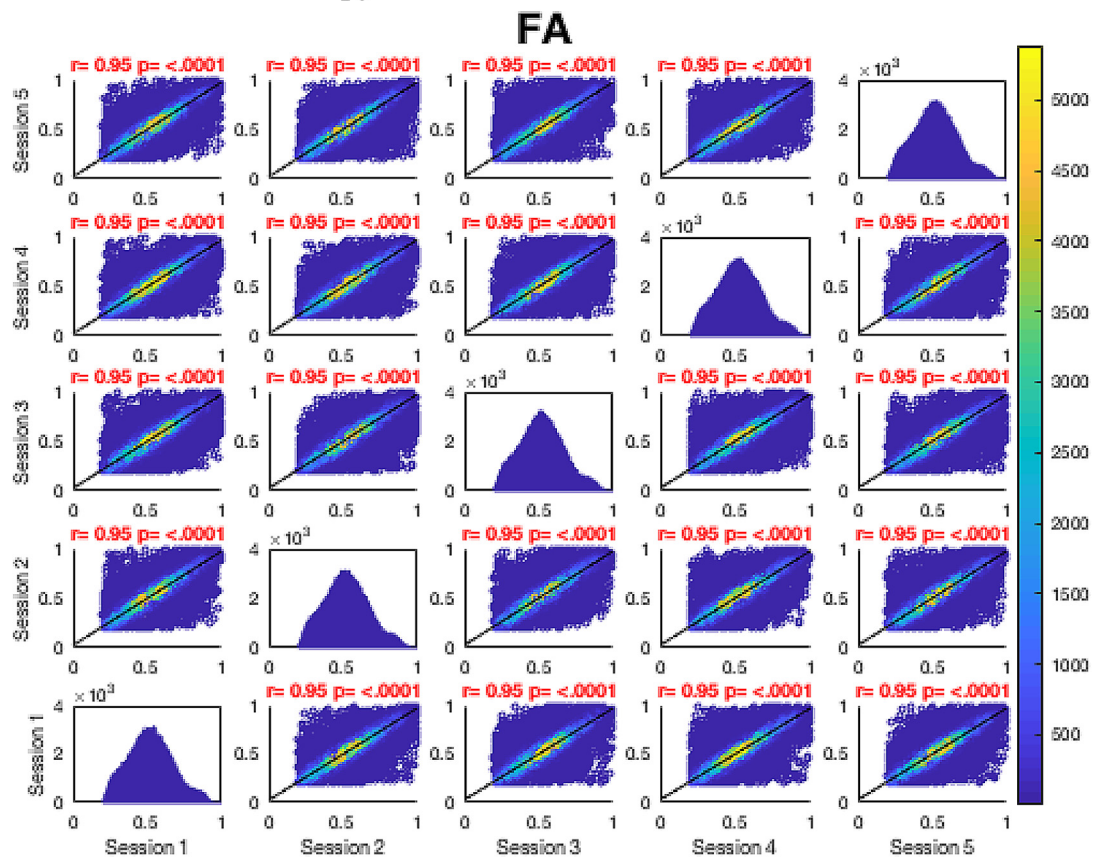


Fig. 4. Voxel-wise whole brain white matter Pearson correlations are presented between individual time points for voxels pooled across all subjects, r = Pearson correlation coefficient, $p < .0001$ for all plots. Fig. 4.1 Fractional Anisotropy (FA), Fig. 4.2 Mean Diffusivity (MD), units = 10^{-3} mm²/s, Fig. 4.3 Radial Diffusivity (RD), units = 10^{-3} mm²/s, Fig. 4.4 Restricted Signal Fraction (RSF), Fig. 4.5 Macromolecular Proton Fraction (MMPF), Fig. 4.6 Myelin Water Fraction (MWF), Fig. 4.7 Longitudinal Relaxation Rate (R₁), units = s⁻¹, Fig. 4.8 Magnetisation Transfer Ratio. For the 2D scatter plots, the colour scale denotes the number of voxels in each joint histogram bin. The univariate histograms on the diagonal show the distributions of voxels of a given metric across all voxels (subject to the inclusion criteria described in 'Methods') for each scan.

of subjects needed to detect a 1% and 5% group difference in each microstructural metric. This was done for all measures and tracts according to means and standard deviations presented in Table 1 (Fig. 4). The minimum n was computed by inputting the percentage change from the mean and standard deviations into G*Power.

(ii) *Group (2) \times Time (2) between-within groups ANOVA (e.g., for showing that there is a difference in the longitudinal evolution of a metric between two groups).* This was estimated across all measures and tracts at small, medium and large effect sizes (Fig. 5). Pearson correlation coefficients were used to account for the correlation amongst repeated measures for sample size estimation (Table 1).

Figure 4.2. Mean Diffusivity, units = $10^{-3} \text{ mm}^2/\text{s}$

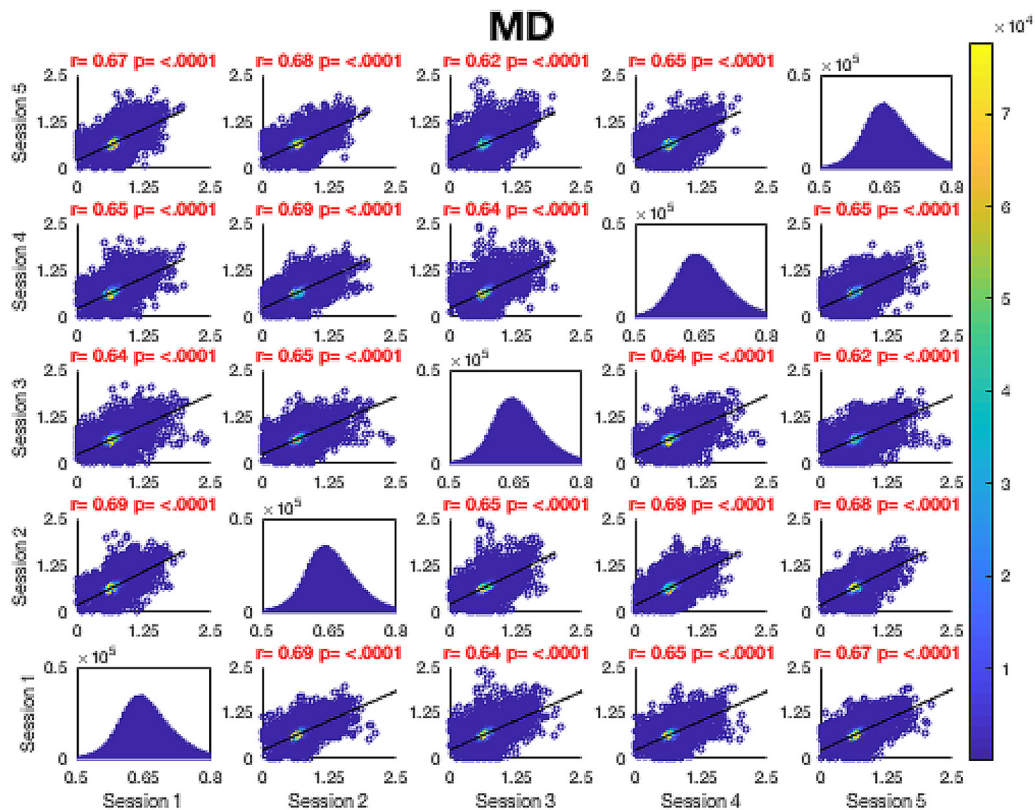


Fig. 4. Continued.

Figure 4.3. Radial Diffusivity, units = $10^{-3} \text{ mm}^2/\text{s}$

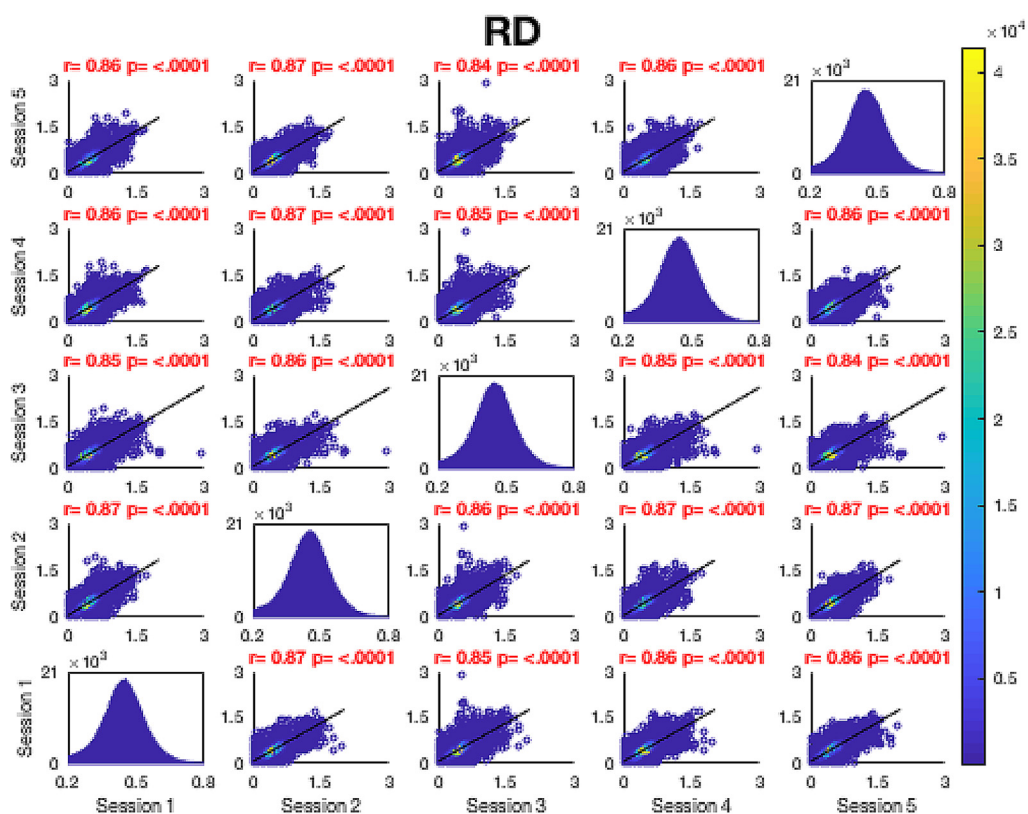


Fig. 4. Continued.

Figure 4.4. Restricted Signal Fraction

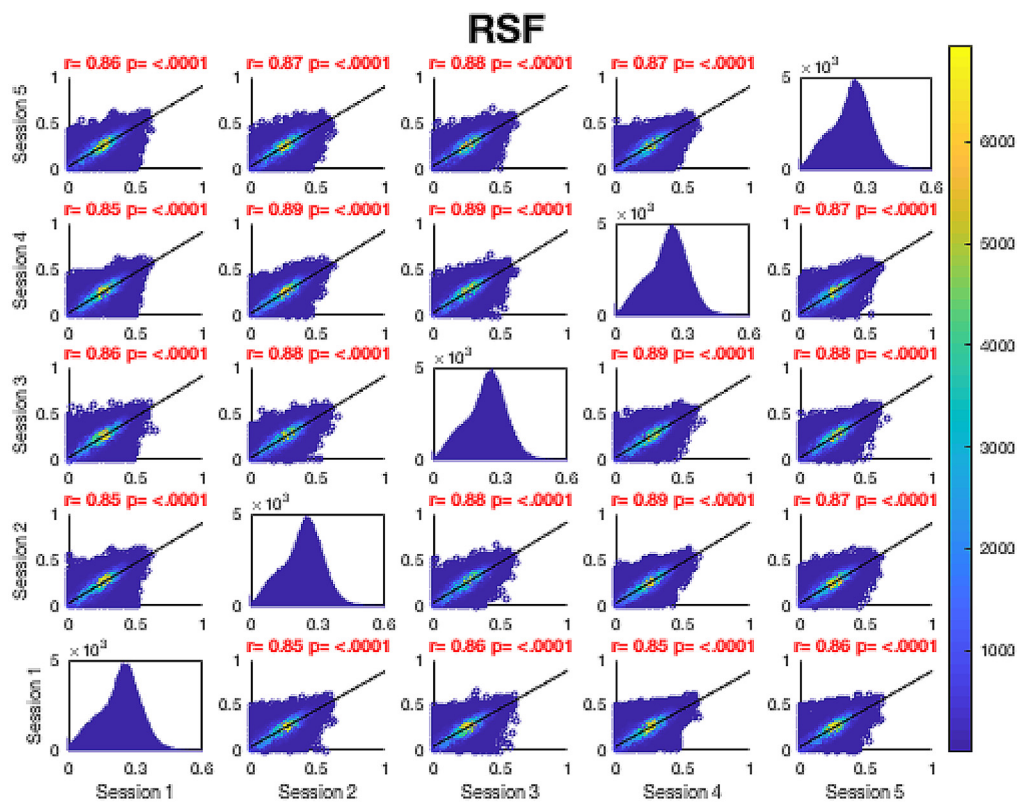


Fig. 4. Continued.

Figure 4.5. Macromolecular Proton Fraction

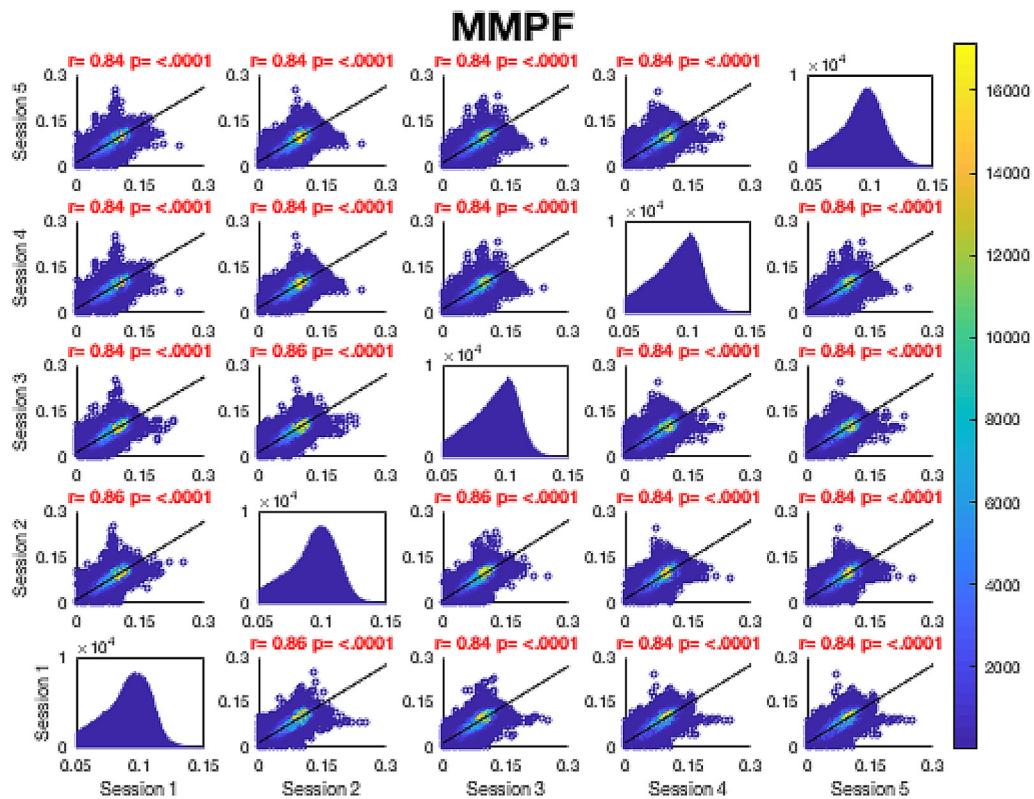


Fig. 4. Continued.

Figure 4.6. Myelin Water Fraction

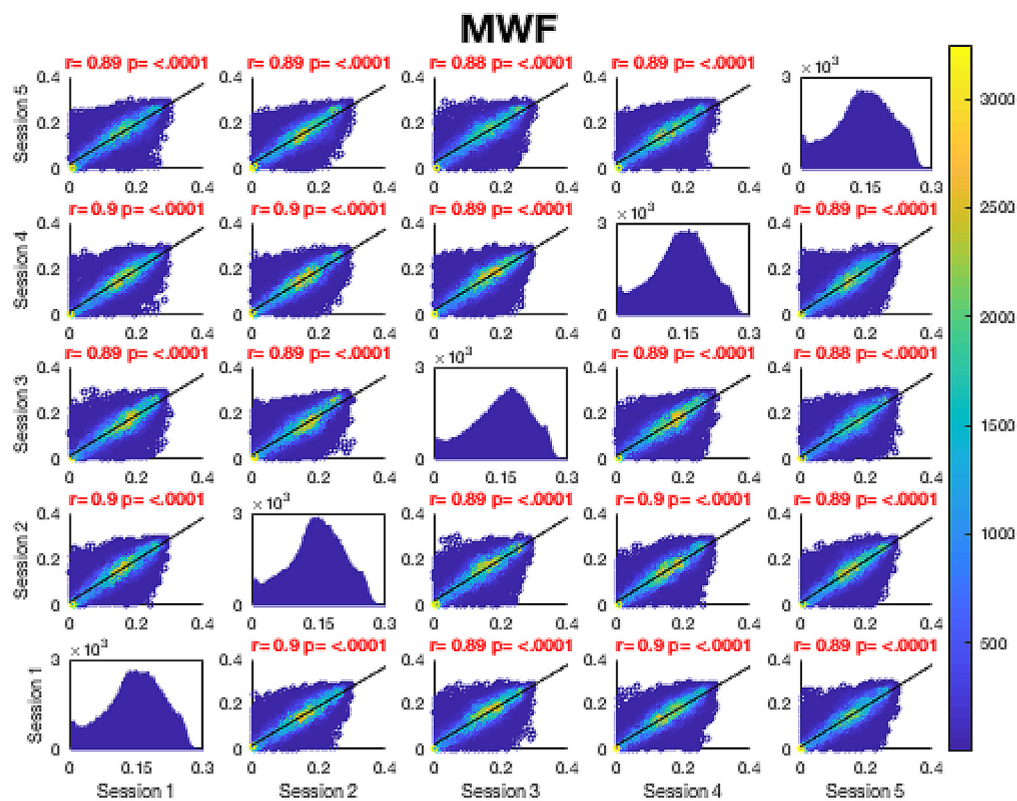


Fig. 4. Continued.

Figure 4.7. Longitudinal Relaxation Rate, units = s^{-1} .

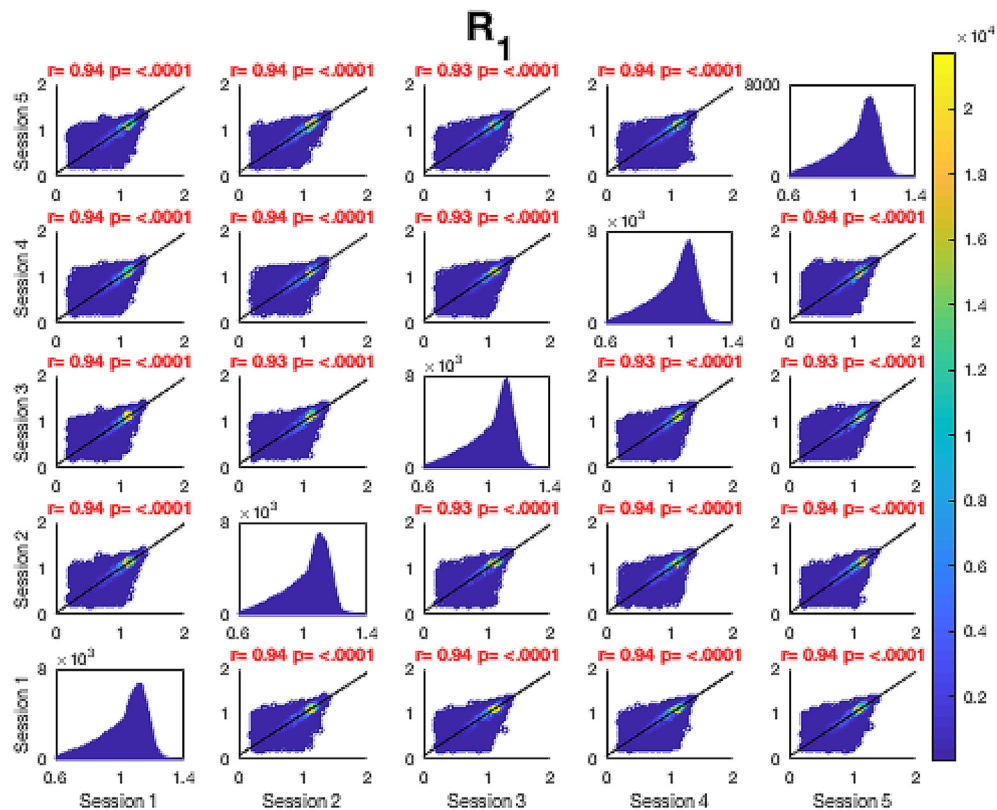


Fig. 4. Continued.

Figure 4.8. Magnetisation Transfer Ratio

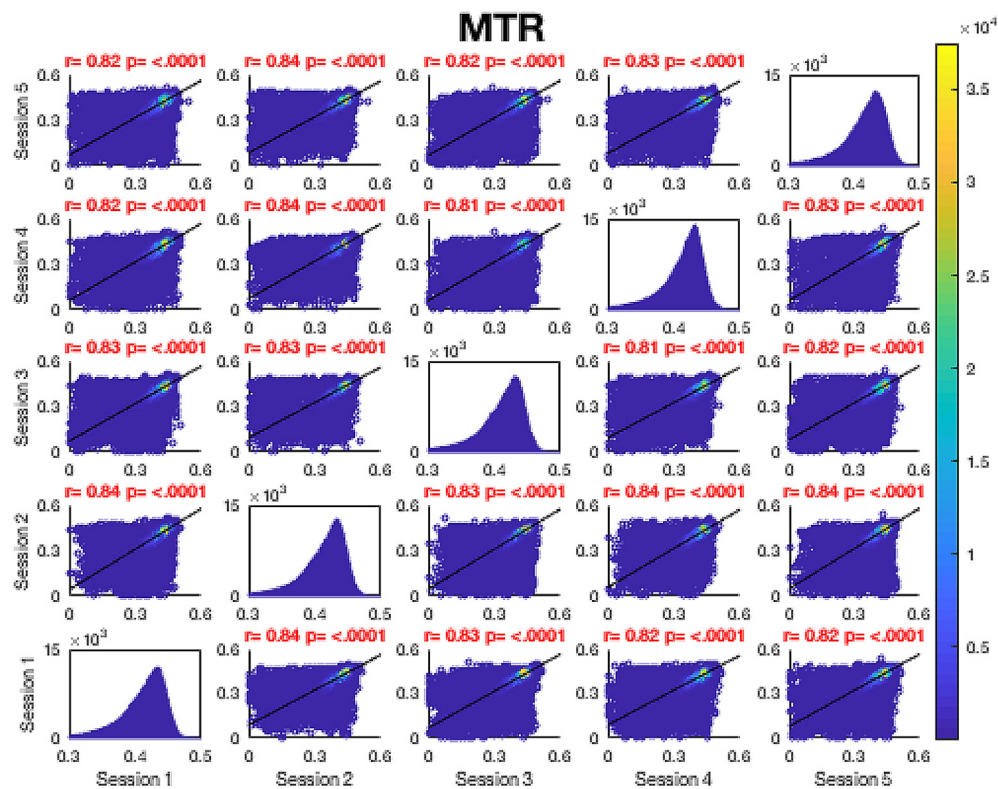


Fig. 4. Continued.

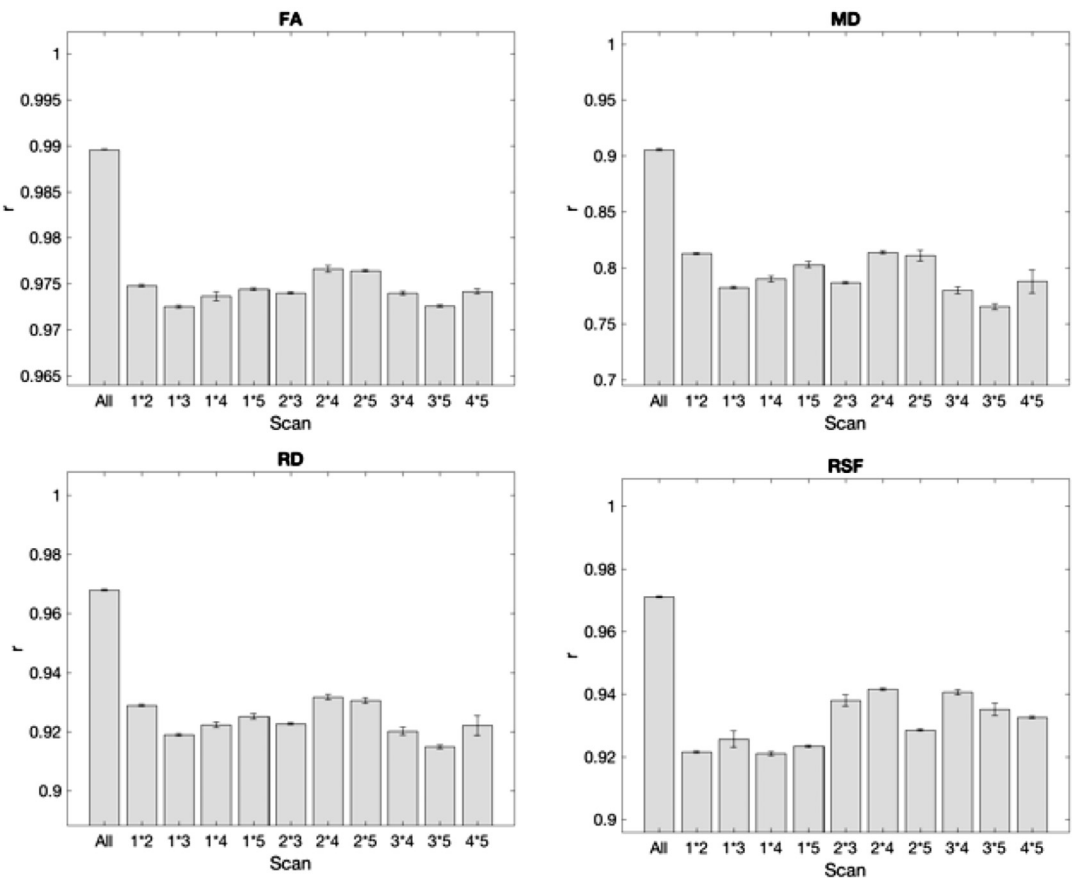


Fig. 5. ICC coefficients for whole brain white matter voxels pooled across all participants. Error bars represent the lower and upper bound confidence intervals.

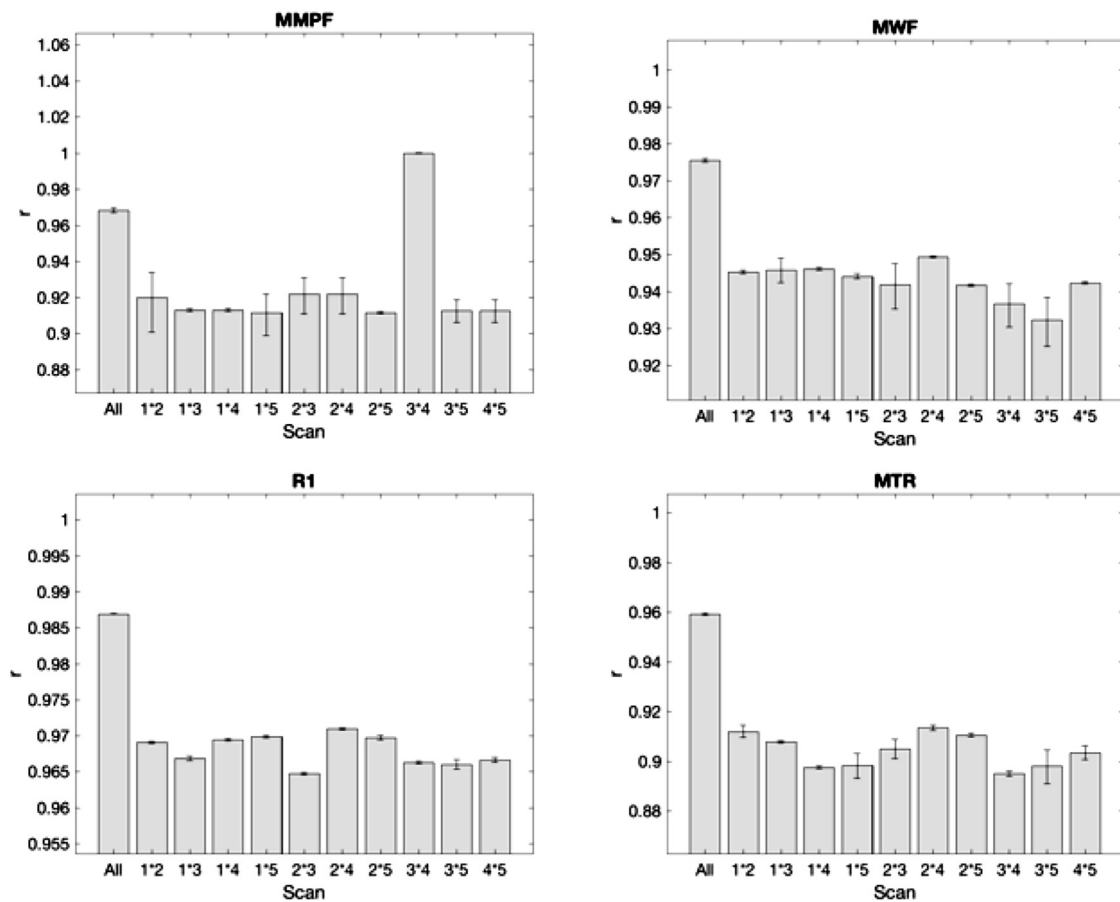


Fig. 5. Continued.

3. Results

Microstructural maps computed for one representative participant are presented in Fig. 1. Fig. 2 shows a typical reconstruction of the fornix, arcuate fasciculus and cortico-spinal tracts, which were successfully dissected bilaterally for each MRI session for each participant. For one participant, calculation of a robust estimate of MWF failed for one session, while for another participant, calculation of the MMPF was not robust in one session. Thus, these values were not included in the analyses.

3.1. Repeatability at the tract level

The coefficients of variation (CV) were overall low, (Table 1), ranging from 0.2 to 4.2%. Intra-class correlations ranged from 0.78 to 0.98 with all demonstrating a high degree of repeatability (Table 1, Fig. 3). Estimated sample sizes for an independent groups *t*-test to measure a 1% and 5% group difference are presented in Fig. 6 and for a 2×2 between-within ANOVA to measure small, medium and large effect sizes in Fig. 7.

The CV values in Table 1 represent the averaged within-subject coefficients of variation. ICC values presented in Table 1 and Fig. 3 represent the two-way mixed effects, absolute agreement with multiple measurements (Mcgraw and Wong, 1996) estimated by the below equation:

$$ICC = \frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$$

MS_R = mean square for rows; MS_E = mean square for error; MS_C = mean square for columns; n = number of subjects (see Koo and Li, 2016, pp. 157 for detailed breakdown of equations for ICC).

3.2. Repeatability at the voxel level

Fig. 4 shows the analysis of repeatability at the voxel-level, for all voxels on the white matter skeleton (see Methods). For each metric, voxelwise whole brain white matter Pearson correlations are presented between individual time points for voxels pooled across all subjects, with a colour scale denoting the number of voxels in each joint histogram bin. Additionally, univariate histograms show the distributions of voxels of a given metric across all voxels.

In terms of the Pearson correlation coefficient, the most reproducible metric is FA with all pair-wise $r > 0.95$ (which is unsurprising given that the FA was used to drive the skeletonisation process). The heatmap representation of the joint histograms show that, despite some considerable scatter, the vast majority of data points lie along the line of identity. For the other metrics: R_1 ($r > 0.93$) shows superb reproducibility. RSF ($r > 0.85$), MMPF ($r > 0.84$), MTR ($r > 0.81$), MWF ($r > 0.88$) and RD ($r > 0.84$) also show good performance. The lowest reproducibility is for MD ($r > 0.62$).

To further ascertain whether there was an effect of time on reproducibility (i.e., do those measurements that are more closely-spaced in time agree better than those spaced further apart?), intra-class correlation coefficients were computed for individual time point pairs across all scan sessions (Fig. 5).

3.3. Demonstration of sample size estimation

Returning to the tract-based estimates, Fig. 6 shows estimated sample sizes for statistical designs required to reach a power of 0.8 and

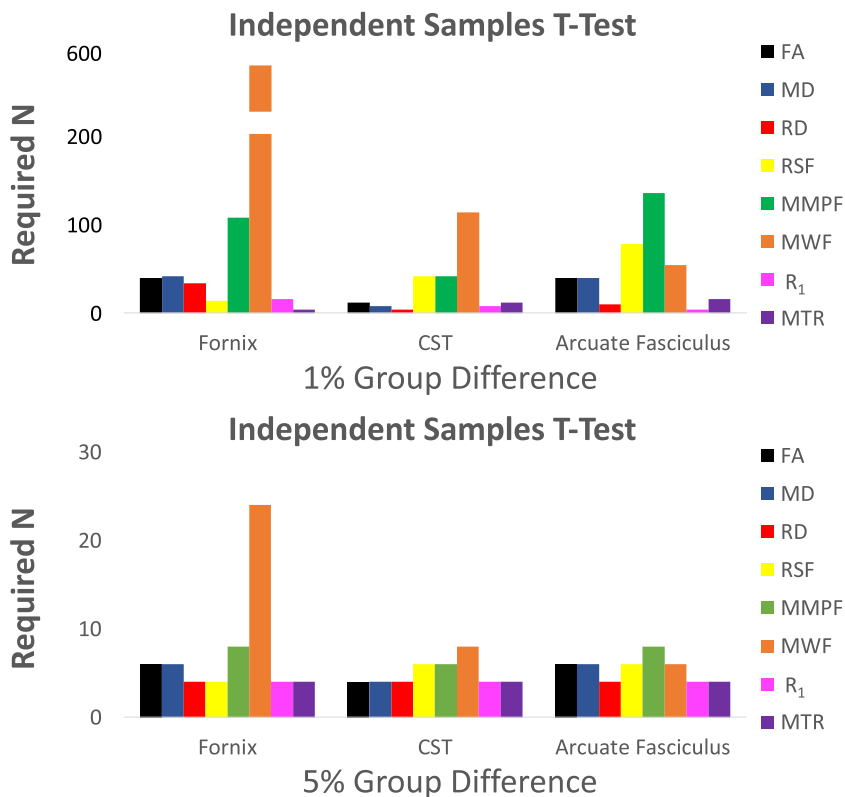


Fig. 6. Estimated sample sizes for statistical designs required to reach a power of 0.8 and significance α of 0.05 for independent groups *t*-test in three white matter tracts across several microstructure measures. Sample sizes were estimated for 1% and 5% group differences according to means and standard deviations presented in Table 1. The standard deviations were assumed to be constant across groups.

significance α of 0.05 for independent groups *t*-test in the fornix, cortico-spinal tract and arcuate fasciculus for the different metrics. Clearly the number of subjects required varies by an order of magnitude depending on which pathway is examined and which metric is used. A similar heterogeneity of required sample sizes is seen when powering for ANOVA analyses (Fig. 7).

4. Discussion

We present this paper as an introduction to MICRA – a multi-variate microstructural dataset collected on an ultra-strong gradient Connectom 3T MRI scanner. We offer the MRI community access to this MRI archive as a test-bed for conducting specialised analyses where access to repeated measures of multi-contrast MRI data may help to inform current and future research. As a demonstration of a possible application of our MRI archive, we explored the reproducibility of microstructural measures, including intra-class correlations and coefficients of variation, across multiple white matter tracts. Additionally, we presented estimates of samples sizes required for an independent groups *t*-test and a Group(2) x Time(2) ANOVA to reach statistical power of 0.8 and significance of $\alpha = 0.05$ for various effect sizes across white matter measures and tracts.

Virtual dissections performed with probabilistic tractography (iFOD2, [Jeurissen et al., 2014](#), MRTrx) demonstrated the fornix, the arcuate fasciculus and cortico-spinal tracts in all six participants and were replicated for all of the five repeated MRI scans. The overall low coefficients of variation within participants and the high correlations among repeated measures suggest a high degree of consistency of microstructure measures across repeated tracts and scans.

Sample size estimations performed for an independent group comparison (*t*-test) across microstructure measures and tracts demonstrated similar patterns of required sample sizes for 1% and 5% increase

changes, with expectedly larger samples required for demonstrations of 1% change. The differences in the standard deviation of measures are reflected in the different sample sizes required. Notably, MWF in the fornix and in the cortico-spinal tract, and MMPF in the Arcuate Fasciculus demonstrated the largest sample sizes required. Conversely, the RSF, MTR, RD and R₁ in the fornix, the MTR, FA, MD, RD and R₁ in the cortico-spinal tract, and the RD, MTR and R₁ in the arcuate fasciculus demonstrated the smallest required sample sizes. Power analyses to estimate sample sizes for a 2 x 2 between-within ANOVA demonstrated that the measures showed a similar pattern for sample size requirement for the fornix and the cortico-spinal tract. In these tracts, measures requiring the smallest sample size were the MWF and R₁. Diffusion measures (FA, RD, MD) and the MMPF required larger sample sizes, whereas the MTR and the RSF required the largest sample sizes to reach a given effect size. In contrast, the arcuate fasciculus demonstrated a pattern in which the diffusion measure required larger sample sizes compared to R₁ and MTR, with MWF requiring the smallest sample size.

To conclude, we present a rich multivariate archive of microstructural MRI data acquired on a Connectom 3T MRI scanner. It is important for researchers to take into consideration that the reproducibility statistics reported here are directly applicable only to scans and analyses that follow conditions unique to the present study conducted on a high gradient Connectom MRI scanner. Although this is unique to the present study, the Connectom-acquired diffusion data offers the highest quality diffusion data available, offering researchers an indication of what might be possible in a 'best case scenario'.

Data from this study demonstrate that microstructure measures derived from multi-shell diffusion, multi-component relaxometry and quantitative magnetisation transfer acquired on an ultra-strong gradient 3T MRI scanner are reliable as demonstrated by low coefficients of variation and high intra-class correlation coefficients across measures and tracts.

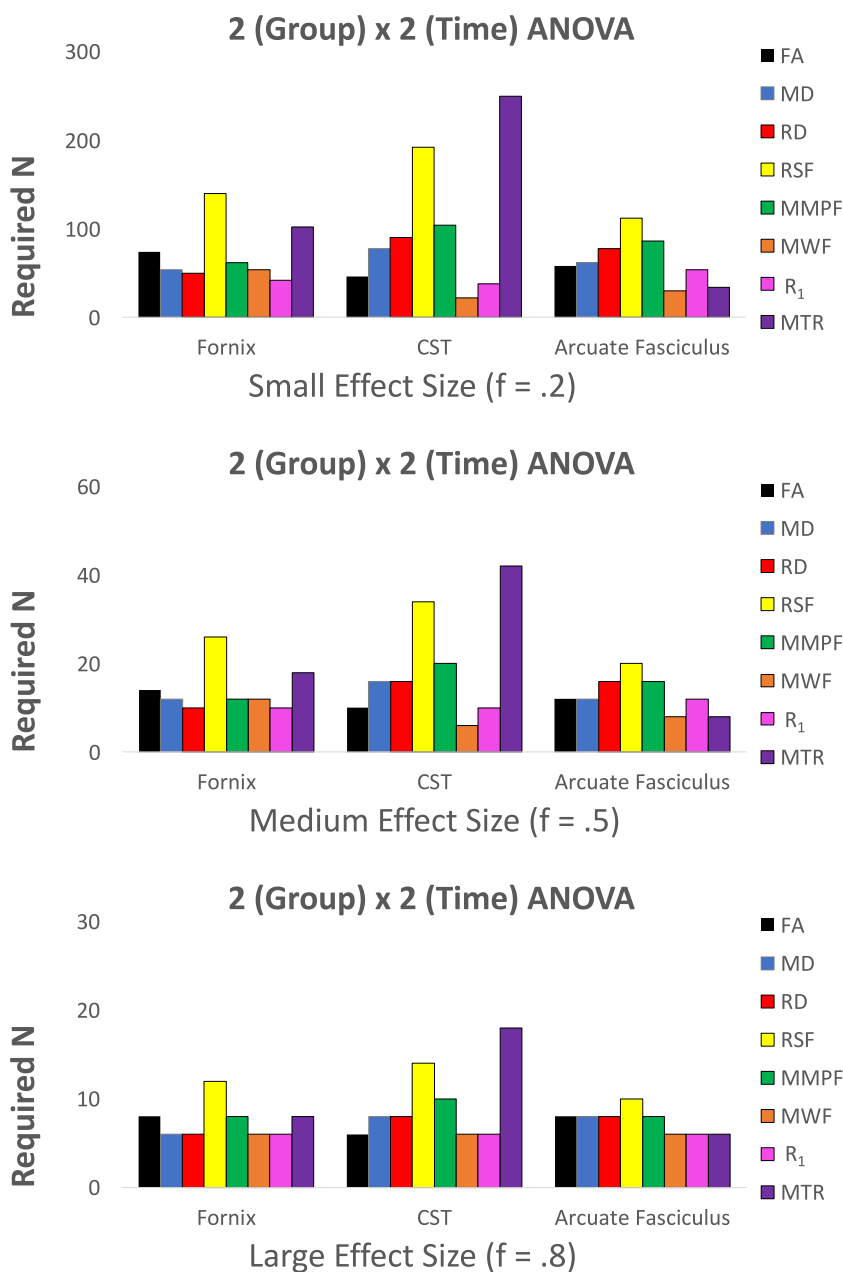


Fig. 7. Estimated sample sizes for statistical designs to reach a power of 0.8 and α of 0.05 in three white matter tracts for a Group (2) x Time Point (2) ANOVA. Pearson correlations between all 5 sessions were averaged by transformation to Fisher's Z (Fisher, 1915) to obtain an average correlation among repeated coefficients for each metric (Table 1). Correlation coefficients were used to estimate required sample sizes for each metric.

Credit authorship contribution statement

Kristin Koller: Conceptualization, Data curation, Formal analysis, Methodology, Writing - original draft, Writing - review & editing. **Umesh Rudrapatna:** Conceptualization, Data curation, Formal analysis, Methodology, Writing - review & editing. **Maxime Chamberland:** Formal analysis, Methodology, Writing - review & editing. **Erika P. Raven:** Data curation, Formal analysis, Funding acquisition, Methodology, Writing - review & editing. **Greg D. Parker:** Formal analysis, Methodology, Writing - review & editing. **Chantal M.W. Tax:** Data curation, Formal analysis, Funding acquisition, Methodology, Writing - review & editing. **Mark Drakesmith:** Conceptualization, Data curation, Methodology, Writing - review & editing. **Fabrizio Fasano:** Methodology, Writing - review & editing. **David Owen:** Software, Writing - review & editing. **Garin Hughes:** Software, Writing - review & editing. **Cyril Charron:** Software, Writing - review & editing. **C John Evans:** Conceptualization, Data curation, Formal analysis, Methodology, Writing - review & editing. **Derek K. Jones:** Conceptualization, Formal analysis,

Funding acquisition, Methodology, Writing - original draft, Writing - review & editing.

Data access

Raw data and processed maps are available on the Open Science Framework - please visit <https://osf.io/z3mkn/>. See Supplementary 1 for outline of data included.

Acknowledgments

The data were acquired at the UK *National Facility for In Vivo MR Imaging of Human Tissue Microstructure* funded by the EPSRC (grant EP/M029778/1), and The Wolfson Foundation. DKJ, KK, GDP, MC and CMW were supported in part by a Wellcome Trust Investigator Award (096646/Z/11/Z) and DKJ and MD by a Wellcome Trust Strategic Award (104943/Z/14/Z). CMWT was supported by a Rubicon grant

(680-50-1527) from the Netherlands Organisation for Scientific Research (NWO). ER was supported by a Marshall-Sherfield Fellowship from the Marshall Aid Commemoration Commission. We would like to sincerely thank Thomas Witzel for his help in establishing the diffusion MRI sequence on the Connectom and Tobias Wood for his help with the QUIT Toolbox. Finally, we would like to thank each participant for their invaluable contribution provided by taking part in this study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2020.117406](https://doi.org/10.1016/j.neuroimage.2020.117406).

References

- Allen, C., Mehler, D.M.A., 2019. Open science challenges, benefits and tips in early career and beyond. *PLOS Biology* 17 (5), e3000246. <https://doi.org/10.1371/journal.pbio.3000246>.
- Andersson, J.L.R., Jenkinson, M., & Smith, S. (2007). *Non-linear registration aka Spatial normalisation FMRIB Technical Report TR07JA2*.
- Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20 (2), 870–888. [https://doi.org/10.1016/S1053-8119\(03\)00336-7](https://doi.org/10.1016/S1053-8119(03)00336-7).
- Andersson, J.L.R., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* 125, 1063–1078. <https://doi.org/10.1016/j.neuroimage.2015.10.019>.
- Assaf, Y., Freidlin, R.Z., Rohde, G.K., Basser, P.J., 2004. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine* 52 (5), 965–978. <https://doi.org/10.1002/mrm.20274>.
- Bells, S., Cercignani, M., Deoni, S., Assaf, Y., 2011. “Tractometry” – comprehensive multi-modal quantitative assessment of white matter along specific tracts. *Proceedings of the International Society for Magnetic Resonance in Medicine* 19 (2009), 678. Retrieved from <http://cds.ismrm.org/protected/11MProceedings/files/678.pdf>.
- Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., ... Milham, M.P., 2010. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America* 107 (10), 4734–4739. <https://doi.org/10.1073/pnas.0911855107>.
- C Wood, T., 2018. QUIT: Quantitative Imaging Tools. *Journal of Open Source Software* 3 (26), 656. <https://doi.org/10.21105/joss.00656>.
- Calamante, F., Tournier, J.D., Jackson, G.D., Connelly, A., 2010. Track-density imaging (TDI): Super-resolution white matter imaging using whole-brain track-density mapping. *NeuroImage* 53 (4), 1233–1243. <https://doi.org/10.1016/j.neuroimage.2010.07.024>.
- Cercignani, M., Alexander, D.C., 2006. Optimal acquisition schemes for in vivo quantitative magnetization transfer MRI. *Magnetic Resonance in Medicine* 56 (4), 803–810. <https://doi.org/10.1002/mrm.21003>.
- Chamberland, M., St-Jean, S., Tax, C.M.W., Jones, D.K., 2019. *Obtaining Representative Core Streamlines for White Matter Tractometry of the Human Brain. Mathematics and Visualization*. Springer, Heidelberg, pp. 359–366.
- Chang, L.C., Jones, D.K., Pierpaoli, C., 2005. RESTORE: Robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine* 53 (5), 1088–1095. <https://doi.org/10.1002/mrm.20426>.
- Deoni, S.C.L., 2009. Transverse relaxation time (T2) mapping in the brain with off-resonance correction using phase-cycled steady-state free precession imaging. *Journal of Magnetic Resonance Imaging* 30 (2), 411–417. <https://doi.org/10.1002/jmri.21849>.
- Deoni, S.C.L., Kolind, S.H., 2015. Investigating the stability of mcDESPOT myelin water fraction values derived using a stochastic region contraction approach. *Magnetic Resonance in Medicine* 73 (1), 161–169. <https://doi.org/10.1002/mrm.25108>.
- Deoni, S.C.L., Matthews, L., Kolind, S.H., 2013. One component? Two components? Three? the effect of including a nonexchanging “free” water component in multicomponent driven equilibrium single pulse observation of T1 and T2. *Magnetic Resonance in Medicine* 70 (1), 147–154. <https://doi.org/10.1002/mrm.24429>.
- Deoni, S.C.L., Rutt, B.K., Arun, T., Pierpaoli, C., Jones, D.K., 2008. Gleaning multicomponent T1 and T2 information from steady-state imaging data. *Magnetic Resonance in Medicine* 60 (6), 1372–1387. <https://doi.org/10.1002/mrm.21704>.
- Ercan, E., Varma, G., Mädler, B., Dimitrov, I.E., Pinho, M.C., Xi, Y., ... Vinogradov, E., 2018. Microstructural correlates of 3D steady-state inhomogeneous magnetization transfer (ihMT) in the human brain white matter assessed by myelin water imaging and diffusion tensor imaging. *Magnetic Resonance in Medicine* 80 (6), 2402–2414. <https://doi.org/10.1002/mrm.27211>.
- Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G., 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41 (4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Fisher, R.A., 1915. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika* 10 (4), 507. <https://doi.org/10.2307/2331838>.
- Friedrich, P., Fraenz, C., Schlüter, C., Ocklenburg, S., Mädler, B., Güntürkün, O., Genç, E., 2020. The Relationship Between Axon Density, Myelination, and Fractional Anisotropy in the Human Corpus Callosum. *Cerebral Cortex* (New York, N.Y.: 1991). <https://doi.org/10.1093/cercor/bhz221>.
- Geeraert, B.L., Lebel, R.M., Lebel, C., 2019. A multiparametric analysis of white matter maturation during late childhood and adolescence. *Human Brain Mapping* 40 (15), 4345–4356. <https://doi.org/10.1002/hbm.24706>.
- Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., ... Andersson, J.L.W.-Minn HCP Consortium, for the W.-M. H., 2013. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80, 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>.
- Grech-Sollars, M., Hales, P.W., Miyazaki, K., Raschke, F., Rodriguez, D., Wilson, M., ... Clark, C.A., 2015. Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. *NMR in Biomedicine* 28 (4), 468–485. <https://doi.org/10.1002/nbm.3269>.
- Hoy, A.R., Koay, C.G., Kecske, S.R., Alexander, A.L., 2014. Optimization of a free water elimination two-compartment model for diffusion tensor imaging. *NeuroImage* 103, 323–333. <https://doi.org/10.1016/j.neuroimage.2014.09.053>.
- 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 (2), 825–841. [https://doi.org/10.1016/S1053-8119\(02\)91132-8](https://doi.org/10.1016/S1053-8119(02)91132-8).
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 5 (2), 143–156. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/10.1016/S1361-8415(01)00036-6).
- Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage* 103, 411–426. <https://doi.org/10.1016/j.neuroimage.2014.07.061>.
- Jones, D.K., Alexander, D.C., Bowtell, R., Cercignani, M., Dell’Acqua, F., McHugh, D.J., ... Tax, C.M.W., 2018, November 15. Microstructural imaging of the human brain with a ‘super-scanner’: 10 key advantages of ultra-strong gradients for diffusion MRI. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2018.05.047>.
- Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine* 42 (3), 515–525. [https://doi.org/10.1002/\(SICI\)1522-2594\(199909\)42:3<515::AID-MRM14>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1522-2594(199909)42:3<515::AID-MRM14>3.0.CO;2-Q).
- Jung, W., Lee, J., Shin, H.G., Nam, Y., Zhang, H., Oh, S.H., Lee, J., 2018, November 15. Whole brain g-ratio mapping using myelin water imaging (MWI) and neurite orientation dispersion and density imaging (NODDI). *NeuroImage*. Academic Press Inc <https://doi.org/10.1016/j.neuroimage.2017.09.053>.
- Kellner, E., Dhital, B., Kiselev, V.G., Reiser, M., 2016. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine* 76 (5), 1574–1581. <https://doi.org/10.1002/mrm.26054>.
- Koo, T.K., Li, M.Y., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 15 (2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.
- Lipp, I., Jones, D.K., Bells, S., Sgarlata, E., Foster, C., Stickland, R., ... Tomassini, V., 2019. Comparing MRI metrics to quantify white matter microstructural damage in multiple sclerosis. *Human Brain Mapping* 40 (10), 2917–2932. <https://doi.org/10.1002/hbm.24568>.
- McGraw, K.O., Wong, S.P., 1996. Forming Inferences about Some Intraclass Correlation Coefficients. *Psychological Methods* 1 (1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>.
- Metzler-Baddeley, C., Foley, S., De Santis, S., Charron, C., Hampshire, A., Caeyenberghs, K., Jones, D.K., 2017. Dynamics of white matter plasticity underlying working memory training: Multimodal evidence from diffusion MRI and relaxometry. *Journal of Cognitive Neuroscience* 29 (9), 1509–1520. <https://doi.org/10.1162/jocn.a.01127>.
- Morris, S.R., Holmes, R.D., Dvorak, A.V., Liu, H., Yoo, Y., Vavasour, I.M., ... Laule, C., 2020. Brain Myelin Water Fraction and Diffusion Tensor Imaging Atlases for 9-10 Year-Old Children. *Journal of Neuroimaging*. <https://doi.org/10.1111/jon.12689>.
- Mougin, O.E., Coxon, R.C., Pitiot, A., Gowland, P.A., 2010. Magnetization transfer phenomenon in the human brain at 7 T. *NeuroImage* 49 (1), 272–281. <https://doi.org/10.1016/j.neuroimage.2009.08.022>.
- Munafo, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie Du Sert, N., ... Ioannidis, J.P.A., 2017, January 10. A manifesto for reproducible science. *Nature Human Behaviour*. Nature Publishing Group <https://doi.org/10.1038/s41562-016-0021>.
- O’Connor, D., Potler, N.V., Kovacs, M., Xu, T., Ai, L., Pellman, J., ... Milham, M.P., 2017. The Healthy Brain Network Serial Scanning Initiative: a resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions. *GigaScience* 6 (2). <https://doi.org/10.1093/gigascience/giw011>.
- Papinutto, N.D., Maule, F., Jovicich, J., 2013. Reproducibility and biases in high field brain diffusion MRI: An evaluation of acquisition and analysis variables. *Magnetic Resonance Imaging* 31 (6), 827–839. <https://doi.org/10.1016/j.mri.2013.03.004>.
- Prčková, V., Rodrigues, P., Puigdelvilló Sanchez, A., Ramos, M., Andorra, M., Martínez-Heras, E., ... Villoslada, P., 2016. Reproducibility of the Structural Connectome Reconstruction across Diffusion Methods. *Journal of Neuroimaging* 26 (1), 46–57. <https://doi.org/10.1111/jon.12298>.
- Ramani, A., Dalton, C., Miller, D.H., Tofts, P.S., Barker, G.J., 2002. Precise estimate of fundamental in-vivo MT parameters in human brain in clinically feasible times. *Magnetic Resonance Imaging* 20 (10), 721–731. [https://doi.org/10.1016/S0730-725X\(02\)00598-2](https://doi.org/10.1016/S0730-725X(02)00598-2).
- Rudrapatna, S.U., Parker, G.D., Roberts, J., Jones, D.K., 2018. Can we correct for interactions between subject motion and gradient-nonlinearity in diffusion {MRI}? In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, 25, p. 1206.

- Rueckert, D., 1999. Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging* 18 (8), 712–721. <https://doi.org/10.1109/42.796284>.
- Setsompop, K., Kimmlingen, R., Eberlein, E., Witzel, T., Cohen-Adad, J., McNab, J.A., ... Wald, L.L., 2013. Pushing the limits of in vivo diffusion MRI for the Human Connectome Project. *NeuroImage* 80, 220–233. <https://doi.org/10.1016/j.neuroimage.2013.05.078>.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Human Brain Mapping* 17 (3), 143–155. <https://doi.org/10.1002/hbm.10062>.
- Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., ... Behrens, T.E.J., 2006. Tract-based spatial statistics: Voxel-wise analysis of multi-subject diffusion data. *NeuroImage* 31 (4), 1487–1505. <https://doi.org/10.1016/j.neuroimage.2006.02.024>.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., ... Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23. <https://doi.org/10.1016/j.neuroimage.2004.07.051>.
- Stikov, N., Boudreau, M., Levesque, I.R., Tardif, C.L., Barral, J.K., Pike, G.B., 2015. On the accuracy of T1 mapping: Searching for common ground. *Magnetic Resonance in Medicine* 73 (2), 514–522. <https://doi.org/10.1002/mrm.25135>.
- Tong, Q., He, H., Gong, T., Li, C., Liang, P., Qian, T., ... Zhong, J., 2019. Reproducibility of multi-shell diffusion tractography on traveling subjects: A multicenter study prospective. *Magnetic Resonance Imaging* 59, 1–9. <https://doi.org/10.1016/j.mri.2019.02.011>.
- Uddin, M.N., Figley, T.D., Solar, K.G., Shatil, A.S., Figley, C.R., 2019. Comparisons between multi-component myelin water fraction, T1w/T2w ratio, and diffusion tensor imaging measures in healthy human brain structures. *Scientific Reports* 9 (1). <https://doi.org/10.1038/s41598-019-39199-x>.
- Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J., Fieremans, E., 2016. Denoising of diffusion MRI using random matrix theory. *NeuroImage* 142, 394–406. <https://doi.org/10.1016/J.NEUROIMAGE.2016.08.016>.
- Vollmar, C., O'Muircheartaigh, J., Barker, G.J., Symms, M.R., Thompson, P., Kumari, V., ... Koepp, M.J., 2010. Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0 T scanners. *NeuroImage* 51 (4), 1384–1394. <https://doi.org/10.1016/J.NEUROIMAGE.2010.03.046>.
- Vos, S.B., Tax, C.M.W., Luijten, P.R., Ourselin, S., Leemans, A., Froeling, M., 2017. The importance of correcting for signal drift in diffusion MRI. *Magnetic Resonance in Medicine* 77 (1), 285–299. <https://doi.org/10.1002/mrm.26124>.
- Wasserthal, J., Neher, P., Maier-Hein, K.H., 2018. TractSeg - Fast and accurate white matter tract segmentation. *NeuroImage* 183, 239–253. <https://doi.org/10.1016/j.neuroimage.2018.07.070>.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm, 20. *IEEE TRANSACTIONS ON MEDICAL IMAGING* Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.200.3832&rep=rep1&type=pdf>.