

# Test-Retest Reliability of Diffusion Tensor Imaging in Huntington's Disease

March 21, 2014 · Biomarkers

## Citation

Cole JH, Farmer RE, Rees EM, Johnson HJ, Frost C, Scahill RI, Hobbs NZ. Test-Retest Reliability of Diffusion Tensor Imaging in Huntington's Disease. PLOS Currents Huntington Disease. 2014 Mar 21 . Edition 1. doi: 10.1371/currents.hd.f19ef63fff962f5cd9c0e88f4844f43b.

[Tweet](#)

## Authors

[James H. Cole](#)

Huntington's Disease Research Group, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK; Computational, Cognitive & Clinical Neuroimaging Laboratory, Department of Medicine, Imperial College London, UK.

[Ruth E. Farmer](#)

Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK.

[Elin M. Rees](#)

Huntington's Disease Research Group, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK.

[Hans J. Johnson](#)

Department of Psychiatry, University of Iowa, Iowa City, Iowa, USA.

[Chris Frost](#)

Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK.

[Rachael I. Scahill](#)

Huntington's Disease Research Group, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK.

[Nicola Z. Hobbs](#)

Huntington's Disease Research Group, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK.

## Abstract

Diffusion tensor imaging (DTI) has shown microstructural abnormalities in patients with Huntington's Disease (HD) and work is underway to characterise how these abnormalities change with disease progression. Using methods that will be applied in longitudinal research, we sought to establish the reliability of DTI in early HD patients and controls. Test-retest reliability, quantified using the intraclass correlation coefficient (ICC), was assessed using region-of-interest (ROI)-based white matter atlas and voxelwise approaches on repeat scan data from 22 participants (10 early HD, 12 controls). T1 data was used to generate further ROIs for analysis in a reduced sample of 18 participants. The results suggest that fractional anisotropy (FA) and other diffusivity metrics are generally highly reliable, with ICCs indicating considerably lower within-subject compared to between-subject variability in both HD patients and controls. Where ICC was low, particularly for the diffusivity

measures in the caudate and putamen, this was partly influenced by outliers. The analysis suggests that the specific DTI methods used here are appropriate for cross-sectional research in HD, and give confidence that they can also be applied longitudinally, although this requires further investigation. An important caveat for DTI studies is that test-retest reliability may not be evenly distributed throughout the brain whereby highly anisotropic white matter regions tended to show lower relative within-subject variability than other white or grey matter regions.

## ***Funding Statement***

This work has been supported by the European Union – PADDINGTON project, contract no. HEALTH-F2-2010-261358. RS is supported by the CHDI/High Q Foundation, a not for profit organization dedicated to finding treatments for Huntington's disease. This work was undertaken at UCLH/UCL, which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme.

## ***Introduction***

Magnetic resonance imaging (MRI) is a valuable tool for investigating the progressive changes caused by neurodegenerative diseases, such as Huntington's Disease (HD). A popular variant of MRI, diffusion weighted imaging (DWI) and the most commonly used analytic model, diffusion tensor imaging (DTI), can offer unique insights into the microstructural properties of both white and grey matter. Accordingly, DTI has been widely used to demonstrate regional neuroanatomical abnormalities in symptomatic HD patients <sup>1,2,3,4,5</sup>, as well as in premanifest HD-gene carriers <sup>6,7,8</sup>, indicating that the technique is sensitive to some of the earliest neuropathological changes in HD.

One major strength of MRI is the ability to collect neuroanatomical data from multiple timepoints on a given sample in order to carry out longitudinal research studies; particularly pertinent in progressive diseases such as HD. Already, volumetric MRI has quantified on-going caudate nucleus atrophy across multiple disease stages <sup>9,10</sup>, and offers utility as potential biomarkers of disease progression. Longitudinal DWI has been performed in HD <sup>11,12,13,14</sup>, but these preliminary studies have been limited in scope and provided inconclusive results (see <sup>15</sup> for review).

Compared to T1- and T2-weighted imaging, DWI is prone to higher levels of image ghosting, susceptibility artefacts, eddy currents and geometric distortions, due to its reliance on single-shot echo-planar imaging <sup>16</sup>. Crucially for movement disorders like HD, DWI is particularly sensitive to subtle bulk motion effects <sup>17</sup>, meaning motion-induced signal loss is greater when compared with volumetric methods. These factors combine to increase the signal variability in DWI, thus it is important to demonstrate adequate test-retest (i.e. scan-rescan) reliability, if DTI is to have utility as a tool for accurately quantifying progressive microstructural changes. If within-subject variability (i.e. scan-rescan variability), influenced by the random occurrence of signal loss and distortion artefacts at different timepoints, is large relative to the between-subject variability at any one timepoint, then sensitivity to meaningful within-subject signal change will be compromised. Metrics such as intraclass correlation coefficient (ICC) can be used to quantify variability in order to provide an indication of the reliability of measurement technique.

Acceptable reliability of DTI metrics *per se* has been demonstrated previously in various settings <sup>18,19,20,21,22,23,24</sup>, including in HD <sup>43</sup>, however, what is apparent from these studies is that estimates of reliability vary considerably. Factors such as the specific image acquisition parameters, scanner characteristics (e.g. field strength, manufacturer), data processing methods or brain region under investigation <sup>22,24,25</sup>, all lead to differing estimates of reliability. Hence the reliability of the specific techniques adopted in a longitudinal DTI study should be established, particularly if those techniques are to ever meet the exacting standards required by clinical trials; the long-term goal of MRI biomarker development. Furthermore, the presence of minor HD-related chorea in early HD patients may mean that such participants are more prone to inducing motion artefacts, reducing the signal-to-noise ratio in their acquired datasets and thus potentially leading to a problematic bias when comparing longitudinal changes with control groups.

The aim of the current study is to quantify the reliability of DTI measures in a sample of early HD patients and controls, specifically using those methods that have demonstrated cross-sectional sensitivity to HD<sup>4</sup> and will be used in on-going longitudinal studies. In addition, reliability will be compared between early HD patients and controls to investigate any potential disease-related group biases that may influence test-retest reliability.

## ***Methods***

### **Participants**

Ten early HD patients and 12 healthy control participants (see Table 1) were scanned at 3T (Siemens) at the Institute of Neurology, University College London. Early HD subjects were required to be within stage I of the disease<sup>26</sup>, defined by a Unified Huntington's Disease Rating Scale (UHDRS) Total Functional Capacity (TFC)  $\geq 11$ , indicating good functional capacity. Control participants were spouses, partners or gene-negative siblings of the early HD subjects. Inclusion criteria included participants being over 18 years of age, free from major psychiatric and concomitant neurological disorders, not currently participating in a clinical trial and no contraindications to MRI. The local ethics committee approved the study and written informed consent was obtained from each participant. Participants were drawn from the larger PADDINGTON study (Pharmacodynamic Approaches to Demonstration of Disease-modification in Huntington's disease by SEN0014196), designed to assess potential biomarkers of HD<sup>4</sup>. This specific subset was solely drawn from the London site where we had access to the scanner and participants were included where additional free time on the scanner was available to allow the repeated diffusion scan.

**Table 1. Demographic characteristics of early HD patients and controls**

Baseline characteristics	Controls		Early HD patients	
<b>N</b>	12		10	
<b>Age (Years)</b>	45.88 (15.69)	23.44-67.87	50.97 (7.68)	42.19-66.46
<b>Sex (male/female)</b>	4/8		3/7	
<b>CAG repeat length</b>			43.3 (2.4)	39 – 46
<b>Disease Burden [age x (cag-35.5)]</b>			388.8 (114.56)	232.6 – 563.4
<b>Total Functional Capacity</b>	13.0 (0.0)	13 – 13	11.5 (1.27)	9 – 13
<b>Total Motor Score</b>	0.125 (0.35)	0 – 1	22.8 (10.79)	7 – 45

Values displayed as mean (standard deviation) followed by range, for continuous variables. Discrete variables show counts of numbers. Disease burden calculated according to the formula by Penney et al., (42). Total functional capacity and total motor score are taken from the Unified Huntington's Disease Rating Scale (UHDRS).

### MRI acquisition

Two diffusion-weighted MRI scans were acquired at 3T for each participant using an EPI sequence with the following parameters: TR = 7600 ms, TE = 84 ms, 65 axial slices of 2 mm thickness, with no inter-slice gaps, acquisition matrix = 96 x 128, in-plane resolution of 2 mm<sup>2</sup>, resulting in isotropic voxels. Diffusion data were acquired in 42 different encoding directions with b = 1000 s/mm<sup>2</sup>, along with 7 b = 0 images. The same protocol was repeated immediately in order to acquire back-to-back datasets for test-retest reliability analysis (i.e. the participant was not removed from the scanner between acquisitions). The scanning session also included collecting a high-resolution T1-weighted MP-RAGE scan for region-of-interest (ROI) segmentation with the following parameters; TR = 2200ms, TE = 2.2ms, flip angle = 10°, FOV = 28cm, matrix size = 256x256, in-plane resolution = 1 mm<sup>2</sup>, slice thickness = 1.0 mm with no inter-slice gap. Visual quality controls assessed the following: compliance with relevant acquisition protocols, minimal artefacts (e.g. movement, intensity) and head positioning.

### Image analysis

#### Pre-processing

For each participant the two DWI scans were randomly assigned into one of two independent pre-processing and statistical analysis streams. Procedures were carried out for each stream separately. This was done to ensure no effects of acquisition order influenced the results. Firstly, diffusion-weighted images were registered to the mean of the seven b0 images to correct for motion and eddy current distortions, and the gradient direction scheme was updated accordingly. Subsequently, a non-linear least-squares method was used to fit the tensor at each voxel. Scalar maps of diffusion metrics such as fractional anisotropy (FA), mean (MD), axial (AD) and radial diffusivity (RD) were then derived from these tensor images. In order to carry out a comprehensive assessment of test-retest reliability, three different approaches were used: a T1 ROI analysis (as per 4) where the analysis was conducted in native diffusion space, an atlas-based automated white matter ROI analysis and a voxelwise analysis.

#### T1 ROI analysis

Four ROIs were defined on the T1 images. For the caudate, corpus callosum and cerebral white matter regions manual delineation was carried out using the MIDAS software package 27. For the putamen the automated BRAINS3 program was used 28. The resulting ROIs were transformed into native diffusion space by first registering the T1 image to the participant's FA image, using an initial affine registration, followed by a non-linear registration, necessary to account for the non-linear distortions found in DWI. This was achieved using Nifty-Reg (<http://sourceforge.net/projects/niftyreg>) for both the affine 29 and non-linear 30 stages. The transformation from T1 to native FA space was then applied to the binary ROI labels using a nearest neighbour interpolation scheme. Registration accuracy for all data was assessed visually to ensure accurate placement of ROIs in diffusion space. The mean FA, MD, AD and RD values across the corpus callosum, cerebral white

matter and bilateral caudate and putamen ROIs was then calculated using FSL (<http://fsl.fmrib.ox.ac.uk>). Four control participants did not have T1-weighted scans and were excluded from this element of the analysis, leaving 10 early HD patients and 8 controls with data for the T1 ROIs.

### Automated atlas-based ROI analysis

Tensor images were converted into DTI-TK format (<http://dti-tk.sourceforge.net>) to run tensor-based registration, a method shown to improve registration accuracy compared with using FA images<sup>31</sup>. Using the standard DTI-TK pipeline, a 'bootstrap' template was defined by an affine registration step to put all subjects into approximately the same space. Each native-space tensor image was non-linearly aligned to this template, using an iterative approach to refine the accuracy of the registration until the difference between successive iterations becomes minimal, based on the Euclidean distance of the tensors<sup>32</sup>. Affine and non-linear transformation parameters were combined to allow the native space tensor images to be warped to common space in a single interpolation step. Once all the images were in a common space, the mean FA map was generated to act as a study specific template. As mentioned above, this was done independently for both processing streams; hence two separate templates were produced.

With all the tensor images aligned to the group template, FA, MD, AD and RD maps were generated for each participant. The next stage was to take the white matter ROIs defined by the ICBM-DTI-81 white matter tract atlas<sup>33</sup>, which is supplied with FSL and contains labels for 48 white matter regions. The 2mm<sup>3</sup> ICBM-DTI-81 atlas image was registered to the group FA template using Nifty-Reg to run an initial affine step followed by a non-linear refinement stage. The resultant transformation was then used to warp the white matter label files to group FA template space, which were finally thresholded at template FA > 0.2 to reduce partial volume effects. Mean FA, MD, AD and RD were then computed across each label region using FSL. This was repeated for the second processing stream and the resultant values were used in the subsequent reliability analyses.

### Statistical analysis

As a simple assessment of agreement, Bland-Altman plots<sup>34</sup> were used to examine variability of DTI metrics within each T1 ROI, with HD patient and control data combined.

A common measure of reliability, the intraclass correlation coefficient (ICC) was used to assess the reliability of the scanning procedure in greater detail, and was calculated for controls and HD patients separately, for each region. Confidence intervals (CIs) at 95% were obtained for ICC values using the delta method. The within- and between-subject variances were also calculated for each measure. ICCs were unadjusted for age and sex in order to avoid making the methods incomparable with previous studies of DTI reliability<sup>22,24</sup>.

### Voxelwise analysis

A similar procedure to the atlas-based analysis was used, with the addition of a within-subject registration step, in order to increase the likelihood of voxelwise correspondence across the brain. Again using DTI-TK to register the tensor images, scans from both processing streams for each participant were co-registered using the initial affine and iterative non-linear steps as detailed above. Once co-registered to an unbiased 'mid-space', the subject means were calculated and then fed into the registration pipeline to define a group template, this time combining scans for both processing streams. The transformations from the separate registration steps were then combined together to allow registration from native space to this combined group space in one interpolation step. The ICC could then be calculated for both FA and MD at each voxel using the *fslmaths* utility in FSL. This registration procedure and subsequent statistical analysis was first completed using the 10 early HD patients and the using data from the 12 control participants.

## Results

### Regional reliability analysis

Test-retest reliability of T1-based and atlas ROIs showed generally high ICCs indicating good levels of reliability (Tables 2-5). This was the case for all diffusion metrics (i.e. FA, MD, RD and AD). However, some

variability in reliability was present across different brain regions.

---

Table 2. ICCs for mean FA values from ICBM-DTI-81 atlas and T1-segmented regions

Region	FA			Controls			
	Early HD	Between-subject $\sigma^2$	Within-subject $\sigma^2$	ICC	Between-subject $\sigma^2$	Within-subject $\sigma^2$	ICC
<b>ICBM-DTI-81 atlas</b>							
Anterior corona radiata Left	14.543	0.334	0.978 (0.923, 0.994)	11.715	0.131	0.989 (0.965, 0.997)	
Anterior corona radiata Right	10.619	0.166	0.985 (0.947, 0.996)	6.887	0.392	0.946 (0.842, 0.983)	
Anterior limb internal capsule Left	3.583	1.045	0.774 (0.444, 0.937)	8.860	1.350	0.868 (0.654, 0.958)	
Anterior limb internal capsule Right	3.972	0.848	0.824 (0.534, 0.95)	8.851	0.410	0.956 (0.869, 0.986)	
Body of corpus callosum	16.494	0.245	0.985 (0.949, 0.996)	3.398	0.473	0.878 (0.676, 0.961)	
Cerebral peduncle Left	5.509	3.818	0.591 (0.206, 0.889)	10.699	2.808	0.792 (0.507, 0.934)	
Cerebral peduncle Right	11.289	5.908	0.656 (0.277, 0.905)	20.161	3.017	0.87 (0.658, 0.959)	
Cingulum (cingulate) Left	10.172	0.433	0.959 (0.865, 0.989)	16.695	0.844	0.952 (0.858, 0.985)	
Cingulum (cingulate) Right	36.891	1.55	0.96 (0.866, 0.989)	12.223	1.222	0.909 (0.748, 0.971)	
Cingulum (Hippocampus) Left	26.133	2.594	0.91 (0.726, 0.975)	14.135	1.155	0.924 (0.786, 0.976)	
Cingulum (Hippocampus) Right	9.467	2.071	0.821 (0.527, 0.949)	4.049	2.598	0.609 (0.253, 0.878)	
Corticospinal tract Left	10.384	3.785	0.733 (0.378, 0.925)	16.761	1.019	0.943 (0.833, 0.982)	
Corticospinal tract Right	11.853	8.939	0.57 (0.187, 0.885)	11.589	5.182	0.691 (0.352, 0.902)	
External capsule Left	6.732	0.369	0.948 (0.831, 0.985)	6.466	0.583	0.917 (0.768, 0.974)	
External capsule Right	9.665	0.37	0.963 (0.877, 0.99)	3.679	0.916	0.801 (0.522, 0.937)	
Fornix cres (stria terminalis) Left	22.871	2.291	0.909 (0.724, 0.974)	5.608	2.091	0.728 (0.405, 0.914)	
Fornix cres (stria terminalis) Right	25.259	2.523	0.909 (0.724, 0.974)	6.746	1.560	0.812 (0.543, 0.940)	
Fornix	19.224	4.661	0.805 (0.497, 0.945)	35.254	5.779	0.859 (0.636, 0.955)	
Genu of corpus callosum	13.518	0.391	0.973 (0.907, 0.992)	6.161	0.266	0.959 (0.876, 0.987)	
Inferior cerebellar peduncle Left	18.695	1.548	0.924 (0.762, 0.978)	2.543	1.377	0.649 (0.298, 0.889)	
Inferior cerebellar peduncle Right	19.428	2.825	0.873 (0.637, 0.964)	3.782	3.376	0.528 (0.174, 0.857)	
Medial lemniscus Left	21.441	3.454	0.872 (0.634, 0.964)	3.792	2.121	0.641 (0.289, 0.887)	
Medial lemniscus Right	19.968	17.091	0.812 (0.228, 0.894)	3.873	9.864	0.382 (0.028, 0.845)	
Posterior tract	26.135	1.898	0.932 (0.786, 0.981)	16.157	1.128	0.935 (0.814, 0.980)	
Post corona radiata Left	12.989	0.805	0.942 (0.813, 0.984)	15.048	0.559	0.964 (0.892, 0.989)	
Post corona radiata Right	7.433	0.323	0.958 (0.862, 0.989)	7.089	0.888	0.889 (0.706, 0.965)	
Posterior limb internal capsule Left	1.979	0.567	0.777 (0.449, 0.937)	3.624	0.404	0.900 (0.726, 0.968)	
Posterior limb internal capsule Right	4.143	0.633	0.867 (0.624, 0.963)	3.228	0.600	0.843 (0.603, 0.950)	
Post thalamic radiation Left	13.326	0.485	0.965 (0.882, 0.990)	13.677	0.253	0.982 (0.944, 0.994)	
Post thalamic radiation Right	10.692	0.36	0.967 (0.890, 0.991)	10.794	0.459	0.959 (0.878, 0.987)	
Retrolecticular internal capsule Left	14.145	0.33	0.977 (0.922, 0.994)	7.065	0.897	0.887 (0.697, 0.964)	
Retrolecticular internal capsule Right	7.444	0.973	0.884 (0.663, 0.967)	8.795	1.254	0.875 (0.670, 0.960)	
Sagittal stratum Left	7.38	0.916	0.890 (0.676, 0.969)	7.771	0.917	0.894 (0.713, 0.966)	
Sagittal stratum Right	6.898	0.637	0.916 (0.741, 0.976)	6.660	0.612	0.916 (0.764, 0.973)	
Splenium of corpus callosum	14.696	0.108	0.993 (0.974, 0.998)	3.788	0.059	0.985 (0.952, 0.995)	
Superior cerebellar peduncle Left	12.546	1.325	0.904 (0.712, 0.973)	5.782	0.714	0.890 (0.703, 0.965)	
Superior cerebellar peduncle Right	5.805	2.503	0.699 (0.330, 0.916)	5.275	2.306	0.696 (0.358, 0.904)	
Superior corona radiata Left	6.476	0.106	0.984 (0.944, 0.995)	12.161	0.204	0.984 (0.949, 0.995)	
Superior corona radiata Right	6.813	0.233	0.967 (0.889, 0.991)	6.943	0.252	0.965 (0.895, 0.989)	
Superior fronto-occipital fasciculus Left	6.427	3.557	0.644 (0.262, 0.902)	14.178	1.469	0.906 (0.741, 0.970)	
Superior fronto-occipital fasciculus Right	33.46	2.985	0.918 (0.748, 0.977)	29.605	11.946	0.712 (0.382, 0.909)	
Superior Long fasciculus Left	7.096	0.153	0.979 (0.928, 0.994)	4.821	0.173	0.965 (0.896, 0.989)	
Superior Long fasciculus Right	7.395	0.286	0.965 (0.884, 0.990)	5.613	0.320	0.946 (0.842, 0.983)	
Uncinate fasciculus Left	7.289	4.292	0.629 (0.246, 0.898)	13.091	1.794	0.880 (0.679, 0.962)	
Uncinate fasciculus Right	7.67	5.624	0.577 (0.193, 0.886)	16.364	3.253	0.834 (0.585, 0.947)	
<b>T1 segmented</b>							
Caudate	13.25	1.253	0.914 (0.736, 0.976)	9.113	0.591	0.939 (0.778, 0.985)	
Putamen	9.875	1.025	0.906 (0.716, 0.974)	10.947	2.199	0.833 (0.507, 0.960)	
White matter	3.559	0.047	0.987 (0.955, 0.996)	1.468	0.077	0.950 (0.815, 0.988)	
Corpus callosum	4.609	0.646	0.877 (0.646, 0.965)	4.932	2.148	0.697 (0.286, 0.929)	

ICC = Intraclass correlation coefficient, shown with 95% confidence intervals in parentheses.  $\sigma^2$  = variance (reported as  $\times 10^4$ ). FA takes values between 0 and 1.





**Table 3. ICC scores for mean MD values from ICBM-DTI-81 atlas and T1-segmented regions**

Region	MD (mm <sup>2</sup> /s)					
	Early HD			Controls		
	Between-subject $\sigma^2$	Within-subject $\sigma^2$	ICC	Between-subject $\sigma^2$	Within-subject $\sigma^2$	ICC
<b>ICBM-DTI-81 atlas</b>						
Anterior corona radiata Left	26.442	0.702	0.974 (0.912, 0.993)	11.316	0.195	0.983 (0.948, 0.995)
Anterior corona radiata Right	26.937	0.609	0.978 (0.924, 0.994)	9.995	0.248	0.976 (0.926, 0.992)
Anterior limb internal capsule Left	13.063	0.438	0.968 (0.891, 0.991)	10.058	0.887	0.919 (0.772, 0.974)
Anterior limb internal capsule Right	68.417	3.234	0.955 (0.852, 0.987)	4.957	1.266	0.797 (0.515, 0.935)
Body of corpus callosum	61.409	1.039	0.983 (0.942, 0.995)	6.922	0.882	0.887 (0.696, 0.964)
Cerebral peduncle Left	41.643	12.127	0.774 (0.444, 0.937)	35.056	13.155	0.727 (0.403, 0.913)
Cerebral peduncle Right	47.989	6.775	0.876 (0.644, 0.965)	89.699	4.026	0.957 (0.872, 0.986)
Cingulum (cingulate) Left	9.637	0.606	0.941 (0.811, 0.983)	7.240	0.848	0.895 (0.715, 0.967)
Cingulum (cingulate) Right	15.163	2.698	0.849 (0.584, 0.957)	6.891	3.630	0.648 (0.297, 0.889)
Cingulum (Hippocampus) Left	54.859	5.804	0.904 (0.712, 0.973)	18.937	0.910	0.954 (0.864, 0.986)
Cingulum (Hippocampus) Right	9.083	4.045	0.692 (0.321, 0.914)	11.525	3.896	0.747 (0.433, 0.920)
Corticospinal tract Left	12.033	5.526	0.685 (0.313, 0.912)	33.520	1.452	0.958 (0.876, 0.987)
Corticospinal tract Right	13.157	3.216	0.804 (0.495, 0.945)	25.712	6.411	0.800 (0.522, 0.936)
External capsule Left	33.534	0.609	0.982 (0.938, 0.995)	7.813	1.401	0.848 (0.612, 0.952)
External capsule Right	24.407	0.882	0.965 (0.883, 0.991)	4.767	1.861	0.719 (0.391, 0.911)
Fornix cres (stria terminalis) Left	82.618	3.438	0.960 (0.868, 0.989)	14.032	2.480	0.850 (0.616, 0.952)
Fornix cres (stria terminalis) Right	139.213	14.200	0.907 (0.720, 0.974)	18.920	2.641	0.878 (0.675, 0.961)
Fornix	237.85	25.696	0.902 (0.707, 0.973)	366.590	61.347	0.857 (0.630, 0.954)
Genu of corpus callosum	32.869	1.160	0.966 (0.886, 0.990)	6.994	0.443	0.940 (0.827, 0.981)
Inferior cerebellar peduncle Left	31.421	2.804	0.918 (0.748, 0.977)	7.854	2.021	0.795 (0.513, 0.935)
Inferior cerebellar peduncle Right	27.993	3.054	0.902 (0.705, 0.972)	11.062	3.159	0.778 (0.483, 0.929)
Medial lemniscus Left	23.719	5.680	0.807 (0.501, 0.946)	9.903	3.431	0.743 (0.426, 0.918)
Medial lemniscus Right	22.725	11.954	0.655 (0.276, 0.905)	16.820	3.511	0.827 (0.572, 0.945)
Pontine tract	26.202	2.727	0.906 (0.716, 0.973)	30.074	1.422	0.955 (0.866, 0.986)
Post corona radiata Left	18.996	0.517	0.974 (0.910, 0.993)	10.309	0.144	0.986 (0.957, 0.996)
Post corona radiata Right	15.214	0.178	0.988 (0.960, 0.997)	8.062	0.613	0.929 (0.798, 0.978)
Posterior limb internal capsule Left	4.308	0.692	0.862 (0.611, 0.961)	5.582	0.665	0.893 (0.711, 0.966)
Posterior limb internal capsule Right	7.543	0.547	0.932 (0.787, 0.981)	3.777	0.691	0.845 (0.607, 0.951)
Post thalamic radiation Left	80.859	1.142	0.986 (0.952, 0.996)	18.529	0.517	0.973 (0.917, 0.991)
Post thalamic radiation Right	58.097	0.168	0.997 (0.990, 0.999)	39.558	1.648	0.960 (0.881, 0.987)
Retrolenticular internal capsule Left	31.405	1.455	0.956 (0.854, 0.988)	10.176	1.206	0.894 (0.712, 0.966)
Retrolenticular internal capsule Right	9.229	0.629	0.936 (0.797, 0.982)	11.271	0.355	0.969 (0.907, 0.990)
Sagittal stratum Left	61.400	2.098	0.967 (0.889, 0.991)	19.546	2.329	0.894 (0.711, 0.966)
Sagittal stratum Right	30.225	0.890	0.971 (0.903, 0.992)	8.229	1.604	0.837 (0.590, 0.948)
Splenium of corpus callosum	86.950	2.454	0.973 (0.907, 0.992)	18.710	1.376	0.932 (0.804, 0.978)
Superior cerebellar peduncle Left	52.724	9.956	0.841 (0.568, 0.955)	29.753	5.799	0.837 (0.590, 0.948)
Superior cerebellar peduncle Right	45.005	6.604	0.872 (0.635, 0.964)	28.428	13.677	0.675 (0.331, 0.897)
Superior corona radiata Left	26.843	0.239	0.991 (0.969, 0.998)	5.634	0.135	0.977 (0.928, 0.993)
Superior corona radiata Right	14.048	0.163	0.989 (0.960, 0.997)	6.000	0.616	0.907 (0.743, 0.970)
Superior fronto-occipital fasciculus Left	54.258	4.331	0.926 (0.769, 0.979)	13.123	1.093	0.923 (0.783, 0.976)
Superior fronto-occipital fasciculus Right	752.008	10.076	0.987 (0.954, 0.996)	22.993	11.770	0.661 (0.314, 0.893)
Superior Long fasciculus Left	15.068	0.191	0.987 (0.956, 0.996)	4.038	0.332	0.924 (0.785, 0.976)
Superior Long fasciculus Right	11.486	0.211	0.982 (0.938, 0.995)	5.617	0.905	0.861 (0.640, 0.956)
Uncinate fasciculus Left	11.163	2.755	0.802 (0.492, 0.944)	10.197	3.186	0.762 (0.457, 0.924)
Uncinate fasciculus Right	27.889	4.902	0.851 (0.587, 0.958)	5.359	1.145	0.824 (0.565, 0.944)
<b>T1 segmented</b>						
Caudate	53.987	10.3	0.840 (0.565, 0.955)	13.788	14.747	0.483 (0.094, 0.894)
Putamen	35.348	1.994	0.947 (0.827, 0.985)	6.388	0.335	0.950 (0.814, 0.988)
White matter	10.947	0.028	0.997 (0.990, 0.999)	6.013	0.102	0.983 (0.933, 0.996)
Corpus callosum	17.210	1.310	0.929 (0.778, 0.980)	9.501	3.043	0.757 (0.372, 0.943)

ICC = Intraclass correlation coefficient, shown with 95% confidence intervals in parentheses.  $\sigma^2$  = variance (reported as  $\times 10^{-10}$ ).



Table 4. ICC scores for mean AD values from ICBM-DTI-81 atlas and T1-segmented regions

Region	AD (mm <sup>2</sup> /s)			Controls		
	Early HD		ICC	Between-subject $\sigma^2$		ICC
	Between-subject $\sigma^2$	Within-subject $\sigma^2$		Between-subject $\sigma^2$	Within-subject $\sigma^2$	
<b>ICBM-DTI-81 atlas</b>						
Anterior corona radiata Left	34.517	0.705	0.980 (0.931, 0.994)	15.443	0.120	0.992 (0.976, 0.998)
Anterior corona radiata Right	33.411	0.632	0.981 (0.936, 0.995)	11.727	0.167	0.986 (0.956, 0.996)
Anterior limb internal capsule Left	10.311	0.943	0.916 (0.743, 0.976)	13.881	1.260	0.917 (0.767, 0.974)
Anterior limb internal capsule Right	58.170	2.873	0.953 (0.846, 0.987)	8.922	1.201	0.881 (0.684, 0.962)
Body of corpus callosum	68.263	0.859	0.988 (0.957, 0.997)	6.887	0.985	0.875 (0.669, 0.960)
Cerebral peduncle Left	37.411	13.336	0.737 (0.385, 0.926)	28.547	12.383	0.697 (0.361, 0.904)
Cerebral peduncle Right	41.698	10.934	0.792 (0.475, 0.942)	80.490	5.277	0.938 (0.822, 0.981)
Cingulum (cingulate) Left	17.121	0.866	0.952 (0.843, 0.986)	13.857	0.968	0.935 (0.812, 0.979)
Cingulum (cingulate) Right	42.360	3.028	0.933 (0.789, 0.981)	10.636	3.883	0.733 (0.411, 0.915)
Cingulum (hippocampus) Left	65.781	6.079	0.915 (0.741, 0.976)	19.260	0.765	0.962 (0.886, 0.988)
Cingulum (hippocampus) Right	13.588	3.275	0.806 (0.499, 0.945)	8.307	4.863	0.631 (0.277, 0.884)
Corticospinal tract Left	15.721	4.669	0.771 (0.438, 0.936)	28.932	0.950	0.968 (0.904, 0.990)
Corticospinal tract Right	10.968	6.857	0.615 (0.231, 0.895)	23.318	6.791	0.774 (0.477, 0.928)
External capsule Left	34.530	0.732	0.979 (0.929, 0.994)	11.350	1.665	0.872 (0.663, 0.959)
External capsule Right	30.312	0.926	0.970 (0.900, 0.992)	6.704	1.651	0.802 (0.525, 0.937)
Fornix cres (stria terminalis) Left	107.033	2.885	0.974 (0.911, 0.993)	17.029	2.812	0.858 (0.634, 0.955)
Fornix cres (stria terminalis) Right	170.419	11.938	0.935 (0.793, 0.982)	22.463	2.881	0.886 (0.695, 0.964)
Fornix	280.675	37.195	0.883 (0.660, 0.967)	403.978	72.456	0.848 (0.612, 0.952)
Genu of corpus callosum	41.355	1.007	0.976 (0.919, 0.993)	10.207	0.399	0.962 (0.887, 0.988)
Inferior cerebellar peduncle Left	36.684	2.057	0.947 (0.828, 0.985)	4.471	1.822	0.710 (0.379, 0.908)
Inferior cerebellar peduncle Right	37.244	2.205	0.944 (0.820, 0.984)	6.292	3.604	0.636 (0.283, 0.886)
Medial lemniscus Left	31.817	4.704	0.871 (0.633, 0.964)	7.474	2.846	0.724 (0.398, 0.912)
Medial lemniscus Right	24.852	14.157	0.637 (0.255, 0.900)	11.327	7.859	0.590 (0.233, 0.873)
Portine tract	28.561	1.753	0.942 (0.814, 0.984)	27.913	1.327	0.955 (0.866, 0.986)
Post corona radiata Left	25.969	0.743	0.972 (0.906, 0.992)	18.864	0.364	0.981 (0.942, 0.994)
Post corona radiata Right	17.554	0.277	0.984 (0.946, 0.996)	10.964	1.068	0.911 (0.753, 0.972)
Posterior limb internal capsule Left	2.757	0.634	0.813 (0.513, 0.947)	5.830	0.576	0.910 (0.750, 0.972)
Posterior limb internal capsule Right	6.461	0.632	0.911 (0.729, 0.975)	4.209	0.589	0.877 (0.675, 0.961)
Post thalamic radiation Left	82.797	1.212	0.986 (0.950, 0.996)	22.157	0.486	0.979 (0.934, 0.993)
Post thalamic radiation Right	65.471	0.320	0.995 (0.983, 0.999)	35.950	1.311	0.965 (0.894, 0.989)
Retrolenticular internal capsule Left	30.320	0.839	0.974 (0.910, 0.993)	10.285	0.918	0.918 (0.770, 0.974)
Retrolenticular internal capsule Right	11.684	0.563	0.954 (0.849, 0.987)	13.256	0.758	0.946 (0.842, 0.983)
Sagittal stratum Left	57.484	2.284	0.962 (0.873, 0.989)	18.403	2.023	0.901 (0.729, 0.969)
Sagittal stratum Right	29.293	0.995	0.967 (0.890, 0.991)	9.436	1.482	0.864 (0.646, 0.957)
Splenium of corpus callosum	85.503	1.652	0.981 (0.935, 0.995)	13.728	0.710	0.951 (0.855, 0.984)
Superior cerebellar peduncle Left	53.160	8.296	0.865 (0.619, 0.962)	29.116	3.931	0.881 (0.683, 0.962)
Superior cerebellar peduncle Right	41.526	7.533	0.846 (0.579, 0.957)	28.840	12.441	0.706 (0.372, 0.907)
Superior corona radiata Left	21.874	0.283	0.987 (0.955, 0.996)	11.505	0.151	0.987 (0.960, 0.995)
Superior corona radiata Right	11.924	0.219	0.982 (0.938, 0.995)	8.040	0.402	0.952 (0.859, 0.985)
Superior fronto-occipital fasciculus Left	41.905	4.250	0.908 (0.721, 0.974)	16.786	1.264	0.930 (0.800, 0.978)
Superior fronto-occipital fasciculus Right	740.662	9.529	0.987 (0.936, 0.996)	32.602	14.908	0.686 (0.345, 0.901)
Superior Long fasciculus Left	17.772	0.229	0.987 (0.956, 0.996)	5.972	0.346	0.945 (0.840, 0.983)
Superior Long fasciculus Right	14.585	0.353	0.976 (0.919, 0.993)	7.792	0.811	0.906 (0.740, 0.970)
Uncinate fasciculus Left	12.355	4.302	0.742 (0.392, 0.928)	13.628	1.510	0.900 (0.727, 0.968)
Uncinate fasciculus Right	26.64	6.510	0.804 (0.495, 0.945)	12.556	2.480	0.835 (0.587, 0.948)
<b>T1 segmented</b>						
Caudate	40.855	9.357	0.814 (0.514, 0.948)	17.083	14.542	0.540 (0.132, 0.901)
Putamen	23.850	2.158	0.917 (0.745, 0.977)	10.808	0.091	0.992 (0.966, 0.998)
White matter	10.896	0.032	0.997 (0.990, 0.999)	4.848	0.066	0.987 (0.945, 0.997)
Corpus callosum	16.002	1.600	0.909 (0.724, 0.974)	6.726	4.363	0.607 (0.187, 0.912)

ICC = Intraclass correlation coefficient, shown with 95% confidence intervals in parentheses.  $\sigma^2$  = variance (reported as  $\times 10^{-10}$ ).



**Table 5. ICC scores for mean RD values from ICBM-DTI-81 atlas and T1-segmented regions**

Region	RD (mm <sup>2</sup> /s)			Controls		
	Between-subject $\sigma^2$	Within-subject $\sigma^2$	ICC	Between-subject $\sigma^2$	Within-subject $\sigma^2$	ICC
<b>ICBM-DTI-81 atlas</b>						
Anterior corona radiata Left	20.881	1.097	0.950 (0.837, 0.986)	16.536	0.610	0.964 (0.893, 0.989)
Anterior corona radiata Right	20.02	0.662	0.968 (0.892, 0.991)	15.495	1.339	0.920 (0.776, 0.975)
Anterior limb internal capsule Left	26.365	0.556	0.979 (0.929, 0.994)	5.116	1.899	0.729 (0.406, 0.914)
Anterior limb internal capsule Right	94.883	4.696	0.953 (0.846, 0.987)	2.235	1.624	0.579 (0.221, 0.869)
Body of corpus callosum	52.012	1.736	0.968 (0.891, 0.991)	10.430	1.011	0.912 (0.754, 0.972)
Cerebral peduncle Left	55.012	11.351	0.829 (0.543, 0.952)	57.487	16.170	0.780 (0.487, 0.930)
Cerebral peduncle Right	63.260	6.169	0.911 (0.729, 0.975)	115.054	8.037	0.935 (0.812, 0.979)
Cingulum (cingulate) Left	7.861	0.391	0.953 (0.845, 0.987)	15.534	1.385	0.918 (0.770, 0.974)
Cingulum (cingulate) Right	10.754	3.332	0.763 (0.426, 0.934)	18.866	4.439	0.810 (0.538, 0.939)
Cingulum (Hippocampus) Left	45.941	7.973	0.852 (0.591, 0.958)	30.529	3.621	0.894 (0.712, 0.966)
Cingulum (Hippocampus) Right	6.444	8.259	0.438 (0.088, 0.863)	22.312	4.709	0.826 (0.569, 0.945)
Corticospinal tract Left	11.500	12.928	0.471 (0.108, 0.867)	50.179	6.264	0.889 (0.701, 0.965)
Corticospinal tract Right	37.609	1.385	0.964 (0.881, 0.990)	39.312	7.472	0.840 (0.597, 0.949)
External capsule Left	34.049	0.713	0.980 (0.930, 0.994)	4.877	1.272	0.793 (0.509, 0.934)
External capsule Right	17.259	1.090	0.941 (0.810, 0.983)	4.192	3.701	0.531 (0.176, 0.857)
Fornix cres (stria terminalis) Left	186.569	14.692	0.927 (0.772, 0.979)	310.713	45.243	0.873 (0.605, 0.960)
Fornix cres (stria terminalis) Right	47.847	7.707	0.861 (0.611, 0.961)	12.659	4.745	0.727 (0.403, 0.913)
Fornix	87.584	25.451	0.775 (0.445, 0.937)	18.968	4.085	0.823 (0.563, 0.944)
Genu of corpus callosum	19.400	2.385	0.891 (0.678, 0.969)	8.161	1.820	0.818 (0.553, 0.942)
Inferior cerebellar peduncle Left	30.184	7.371	0.804 (0.495, 0.945)	26.136	4.148	0.863 (0.644, 0.956)
Inferior cerebellar peduncle Right	20.275	9.009	0.692 (0.322, 0.914)	33.787	6.326	0.842 (0.601, 0.950)
Medial lemniscus Left	22.343	11.173	0.667 (0.289, 0.908)	22.377	8.483	0.725 (0.400, 0.913)
Medial lemniscus Right	40.284	18.744	0.682 (0.309, 0.912)	36.803	10.301	0.781 (0.489, 0.930)
Pontine tract	36.048	6.429	0.849 (0.583, 0.957)	45.794	2.408	0.950 (0.853, 0.984)
Post corona radiata Left	14.604	1.417	0.912 (0.790, 0.975)	11.357	0.610	0.949 (0.850, 0.984)
Post corona radiata Right	17.075	0.446	0.975 (0.913, 0.993)	10.101	0.549	0.948 (0.849, 0.984)
Posterior limb internal capsule Left	14.552	1.228	0.922 (0.759, 0.978)	8.181	1.320	0.861 (0.640, 0.956)
Posterior limb internal capsule Right	14.146	1.039	0.932 (0.784, 0.981)	7.600	2.512	0.752 (0.440, 0.921)
Post thalamic radiation Left	83.201	1.434	0.983 (0.941, 0.995)	27.310	0.914	0.968 (0.902, 0.990)
Post thalamic radiation Right	56.126	0.415	0.993 (0.974, 0.998)	65.956	4.002	0.943 (0.833, 0.982)
Retrolenticular internal capsule Left	41.716	4.251	0.908 (0.720, 0.974)	26.427	3.963	0.870 (0.658, 0.959)
Retrolenticular internal capsule Right	14.038	2.600	0.844 (0.573, 0.956)	15.150	1.960	0.885 (0.693, 0.964)
Sagittal stratum Left	72.078	2.000	0.973 (0.908, 0.992)	28.940	4.783	0.862 (0.642, 0.956)
Sagittal stratum Right	34.265	2.110	0.942 (0.814, 0.984)	19.789	2.492	0.888 (0.699, 0.964)
Splenium of corpus callosum	92.998	4.607	0.953 (0.846, 0.987)	37.418	3.561	0.913 (0.757, 0.972)
Superior cerebellar peduncle Left	55.447	18.223	0.753 (0.409, 0.931)	33.724	13.131	0.720 (0.392, 0.911)
Superior cerebellar peduncle Right	54.641	6.677	0.891 (0.679, 0.969)	27.011	18.825	0.589 (0.231, 0.872)
Superior corona radiata Left	49.954	0.441	0.991 (0.969, 0.998)	7.217	0.439	0.943 (0.833, 0.982)
Superior corona radiata Right	30.796	0.306	0.990 (0.966, 0.997)	9.310	1.472	0.863 (0.645, 0.957)
Superior fronto-occipital fasciculus Left	89.146	9.292	0.906 (0.715, 0.973)	16.854	3.337	0.835 (0.586, 0.947)
Superior fronto-occipital fasciculus Right	802.01	14.854	0.982 (0.937, 0.995)	13.263	14.255	0.482 (0.136, 0.846)
Superior Long fasciculus Left	11.903	0.303	0.975 (0.915, 0.993)	5.691	0.474	0.923 (0.783, 0.976)
Superior Long fasciculus Right	9.706	0.160	0.984 (0.944, 0.995)	9.221	1.307	0.876 (0.671, 0.961)
Uncinate fasciculus Left	20.291	4.319	0.824 (0.534, 0.951)	20.502	13.649	0.600 (0.243, 0.875)
Uncinate fasciculus Right	31.987	9.398	0.773 (0.441, 0.936)	14.987	4.851	0.755 (0.446, 0.922)
<b>T1 segmented</b>						
Caudate	88.405	13.67	0.866 (0.621, 0.962)	9.376	15.571	0.376 (0.042, 0.893)
Putamen	70.692	2.653	0.964 (0.879, 0.990)	1.987	3.218	0.382 (0.044, 0.892)
White matter	9.624	0.105	0.989 (0.962, 0.997)	9.972	0.347	0.966 (0.870, 0.992)
Corpus callosum	32.867	1.626	0.953 (0.846, 0.987)	34.491	2.862	0.923 (0.730, 0.982)

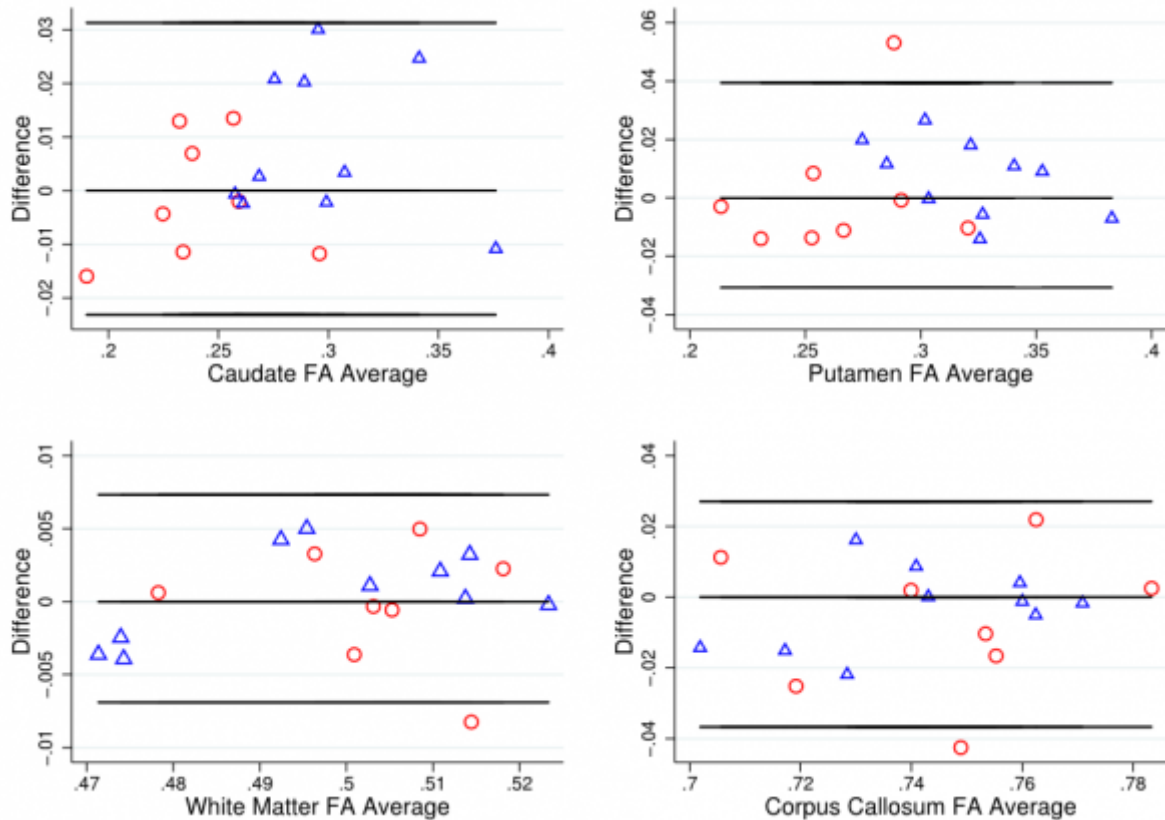
ICC = intraclass correlation coefficient, shown with 95% confidence intervals in parentheses.  $\sigma^2$  = variance (reported as  $\times 10^{-10}$ ).

For FA in the controls, 38 (79%) atlas ROIs had an ICC of 0.8 or above, with the majority of these being  $> 0.9$ . Exceptions to this included the corticospinal tract and cerebellar ROIs. This pattern was reflected in the early HD patients, with 41 (85%) ROIs having ICCs  $> 0.8$ . Generally, between-subject variance was larger in the HD group than controls, as would be expected. It is also worth noting that the within-subject variances are relatively similar for controls and HD in many regions. This implies that, in general, differences in ICC observed between controls and HD are driven by the between-subject variation rather than by the scanning technique being less reliable in one group than the other.

For MD 37 (77%) atlas ROIs had ICCs above 0.8 in the controls; the corresponding number was 44 (92%) in the early HD patients. As with FA, any between group differences in ICC were accompanied by wide 95% CIs around each group estimate, demonstrating the imprecision in the estimates. Results from AD were similar, where 33 (69%) atlas ROIs had ICCs  $> 0.8$ , and RD with 37 (77%) ROIs having ICCs  $> 0.8$ . In controls mean ICCs were very similar across metrics (FA mean ICC = 0.851; MD = 0.854; AD = 0.811; RD = 0.857). However, a few ROIs did show discrepancies between metrics, such as the right anterior limb of internal capsule and right external capsule, which exhibited substantially lower ICCs in AD compared to the other metrics. This is driven by the much smaller between-subject variation for AD than other for metrics in the right anterior limb of internal capsule, and by an increased within-subject variability for the external capsule. Examination of scatter plots of the data identified two outliers that may have contributed to this.

For the T1 ROIs, FA reliability was also high, with the caudate, putamen and whole-brain white matter regions having ICCs  $> 0.8$  in controls and early HD patients. The corpus callosum region had a lower ICC in the controls (ICC = 0.697) though the ICC was high in the early HD patients (ICC = 0.877). For MD, there was considerable regional variability in reliability, with very high ICCs for the putamen and white matter (ICC  $> 0.95$ ), an intermediate value for the corpus callosum (ICC = 0.76) and a low value for the caudate (ICC = 0.48). Furthermore, the ICC in the early HD patients was reasonably high for caudate MD (ICC = 0.84). For AD, both the caudate and the putamen had low reliability in the controls (ICCs  $< 0.4$ ), though for RD the putamen values were very high (ICC = 0.99). The corpus callosum ROI showed the converse pattern, with a very high ICC for AD (ICC = 0.91) and a low one for RD (ICC = 0.61). The white matter ROI showed very high reliability (ICCs  $> 0.9$ ) across the board.

To put these divergent results from the T1 ROIs into context, the Bland-Altman (BA) plots were considered for FA (Figure 1) and MD (Figure 2). These generally suggest good agreement, with the differences tending to lie within a small range. The exceptions to this were MD in the caudate, and FA and MD in the corpus callosum. In these cases, there was a suggestion of deviation from exact reproducibility in the combined HD patient and control group. However, the BA plots do not suggest that the difference between scans is dependent on the magnitude of the measurement in question, and indicated the possibility of outliers within the data.

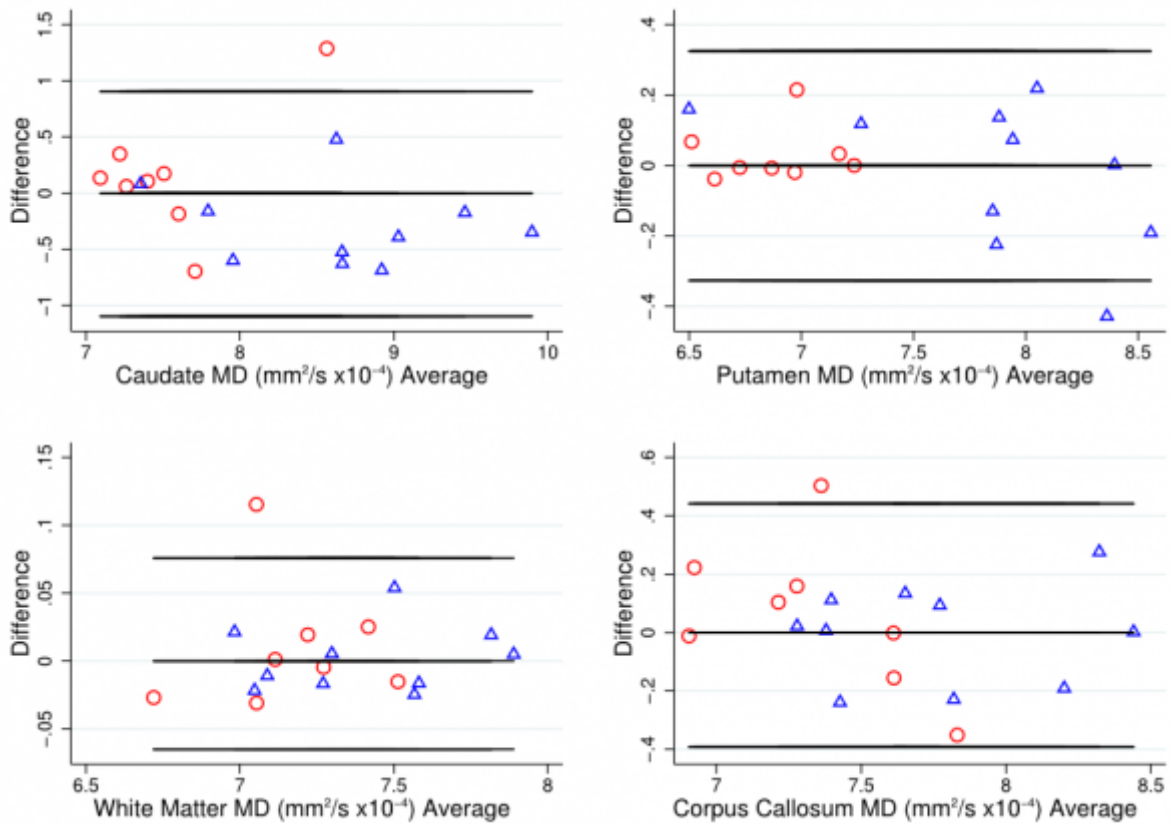


**Fig. 1: Bland-Altman plot for FA values**

Bland Altman plots of fractional anisotropy (FA) values to visually assess agreement, systematic bias and proportional bias in scanning technique for T1 ROIs (caudate, putamen, white matter and corpus callosum), for early HD patients (blue triangles) and controls (red circles). FA is a relative value derived from the diffusion tensor, where 0 indicates perfectly isotropic tensor dimensions (i.e. a sphere) and 1 indicates the maximum theoretical level of anisotropy.



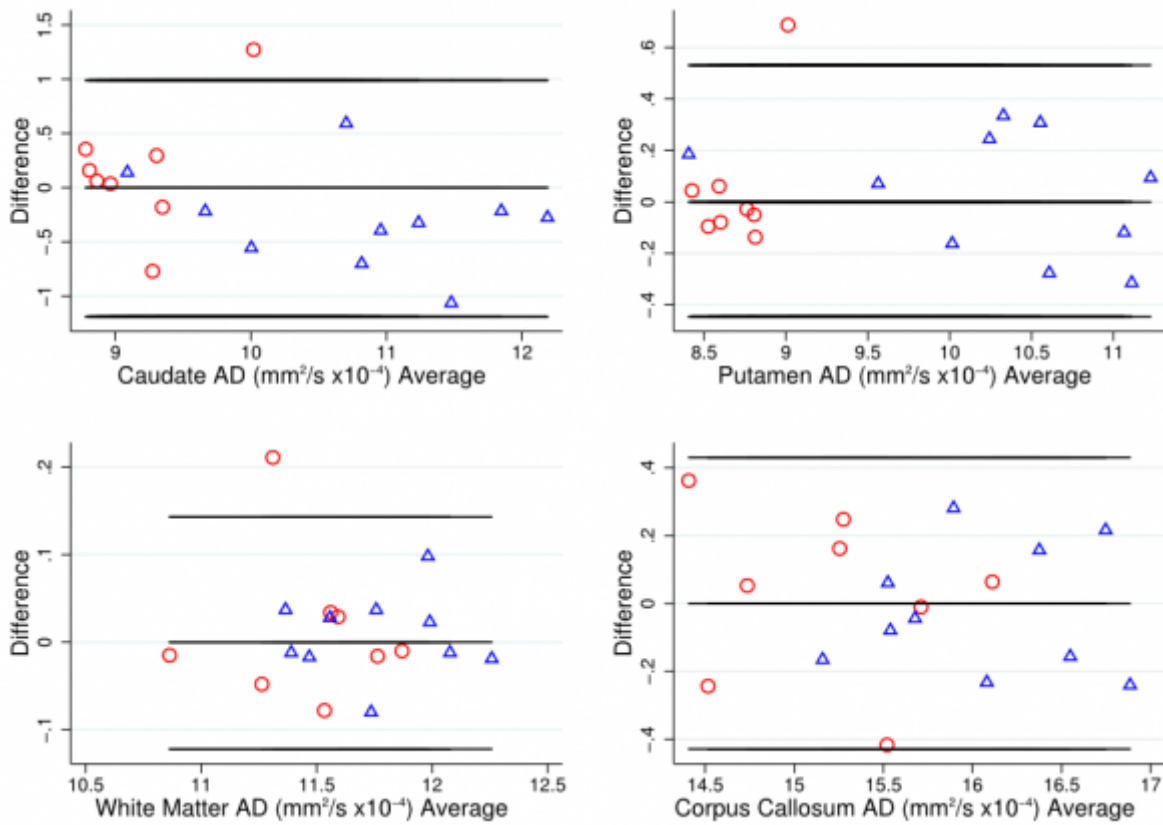




**Fig. 2: Bland-Altman plot for MD values**

Bland Altman plots of mean diffusivity (MD) values to visually assess agreement, systematic bias and proportional bias in scanning technique for T1 ROIs (caudate, putamen, white matter and corpus callosum), for early HD patients (blue triangles) and controls (red circles).

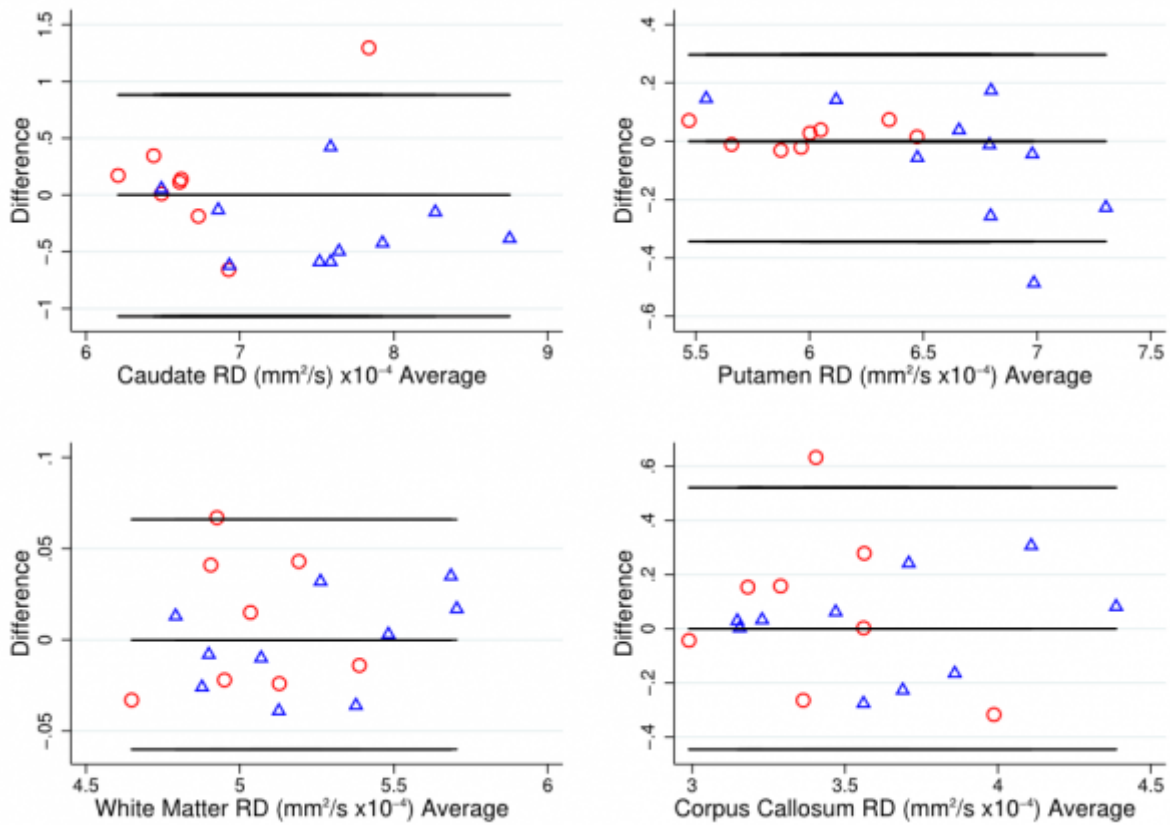




**Fig. 3: Bland-Altman plot for AD values**

Bland Altman plots of axial diffusivity (AD) values to visually assess agreement, systematic bias and proportional bias in scanning technique for T1 ROIs (caudate, putamen, white matter and corpus callosum), for early HD patients (blue triangles) and controls (red circles).





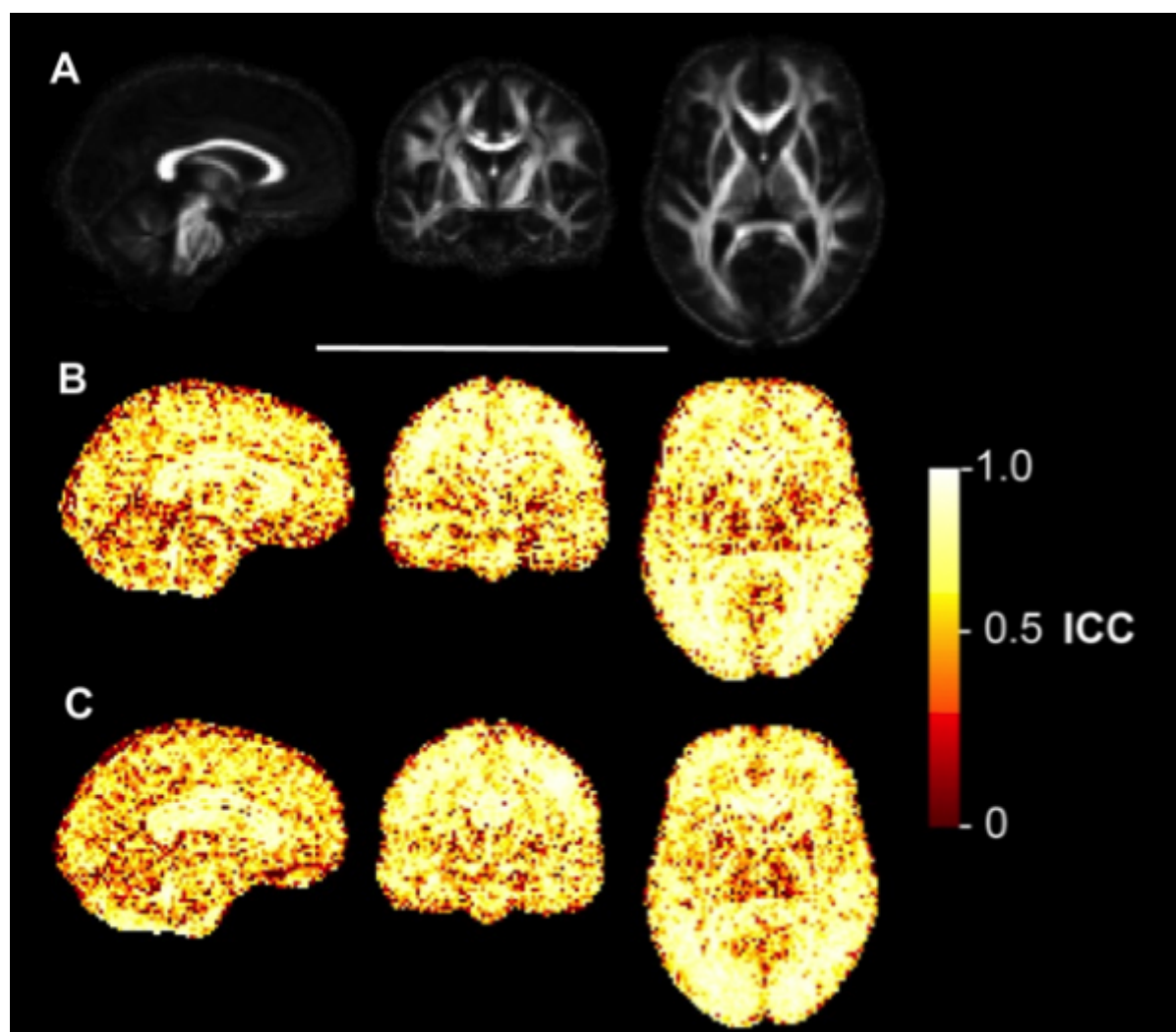
**Fig. 4: Bland-Altman plot for RD values**

Bland Altman plots of radial diffusivity (RD) values to visually assess agreement, systematic bias and proportional bias in scanning technique for T1 ROIs (caudate, putamen, white matter and corpus callosum), for early HD patients (blue triangles) and controls (red circles).

### **Voxelwise reliability analysis**

The ICC maps for FA (Figure 3), MD (Figure 4), AD (Figure 5) and RD (Figure 6) indicated that the reliability was generally high across the brain, with the exception of areas inferior to the lateral ventricles. There is also a degree of noise evident in these maps, likely due to residual registration error between voxels. The group maps appeared qualitatively similar between controls and early HD patients. FA and MD also seemed to generate similar patterns of voxelwise reliability, although the diffusivity metrics (i.e. MD, AD, RD) tend to show consistently higher ICC scores with less regional variability than FA. Lower ICCs were evident in the basal ganglia regions for all four measures.

---

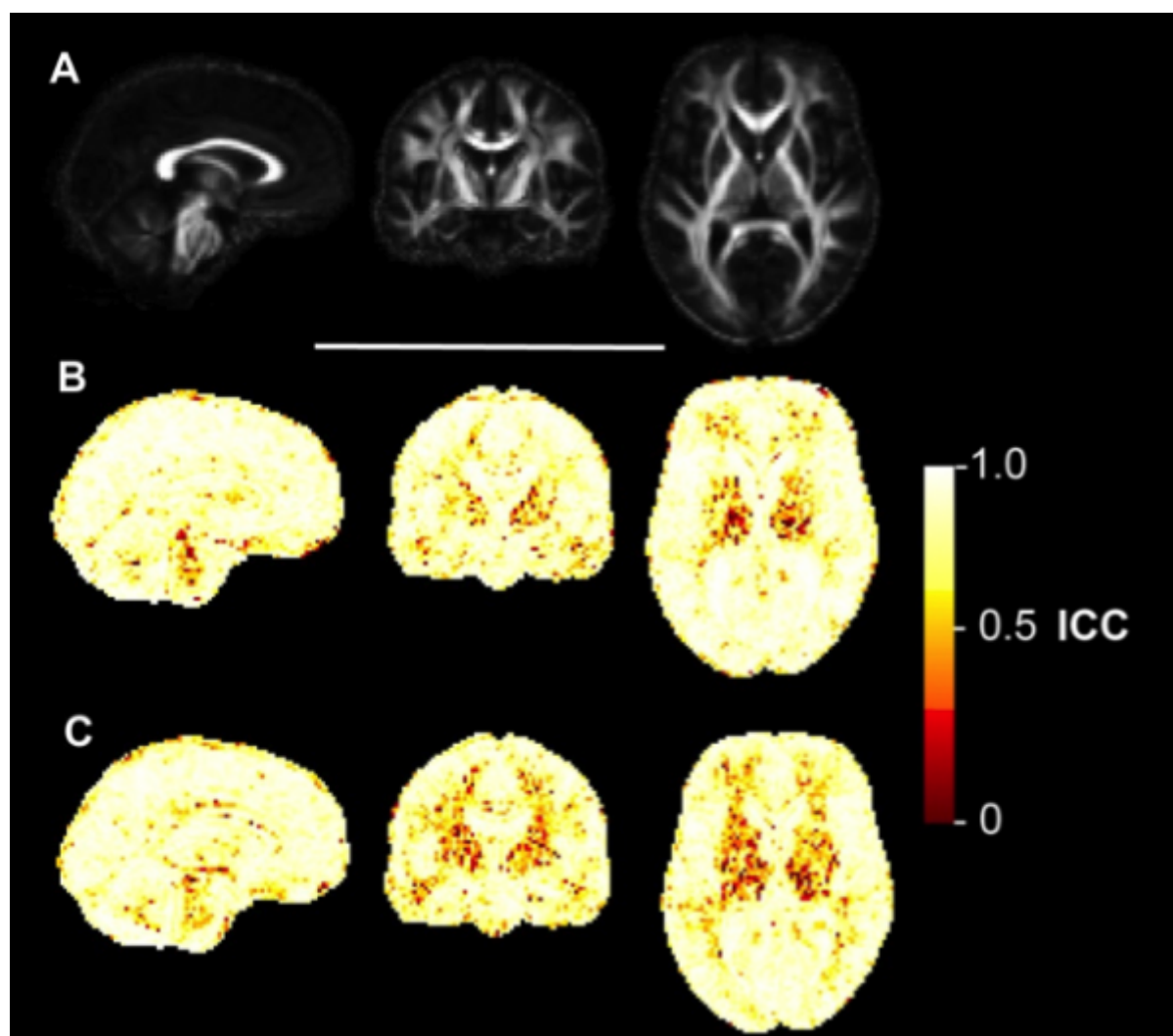


**Fig. 5: Voxelwise ICC maps of FA**

Voxelwise distribution of reliability metrics for fractional anisotropy (FA). Panel A) shows sagittal, coronal and axial slices of the mean FA image created during the image processing, included for anatomical reference. B) Equivalent three slices for the intraclass correlation coefficient (ICC) of FA in early HD patients. Higher values reflect higher test-retest reliability. C) ICC of FA in healthy controls.



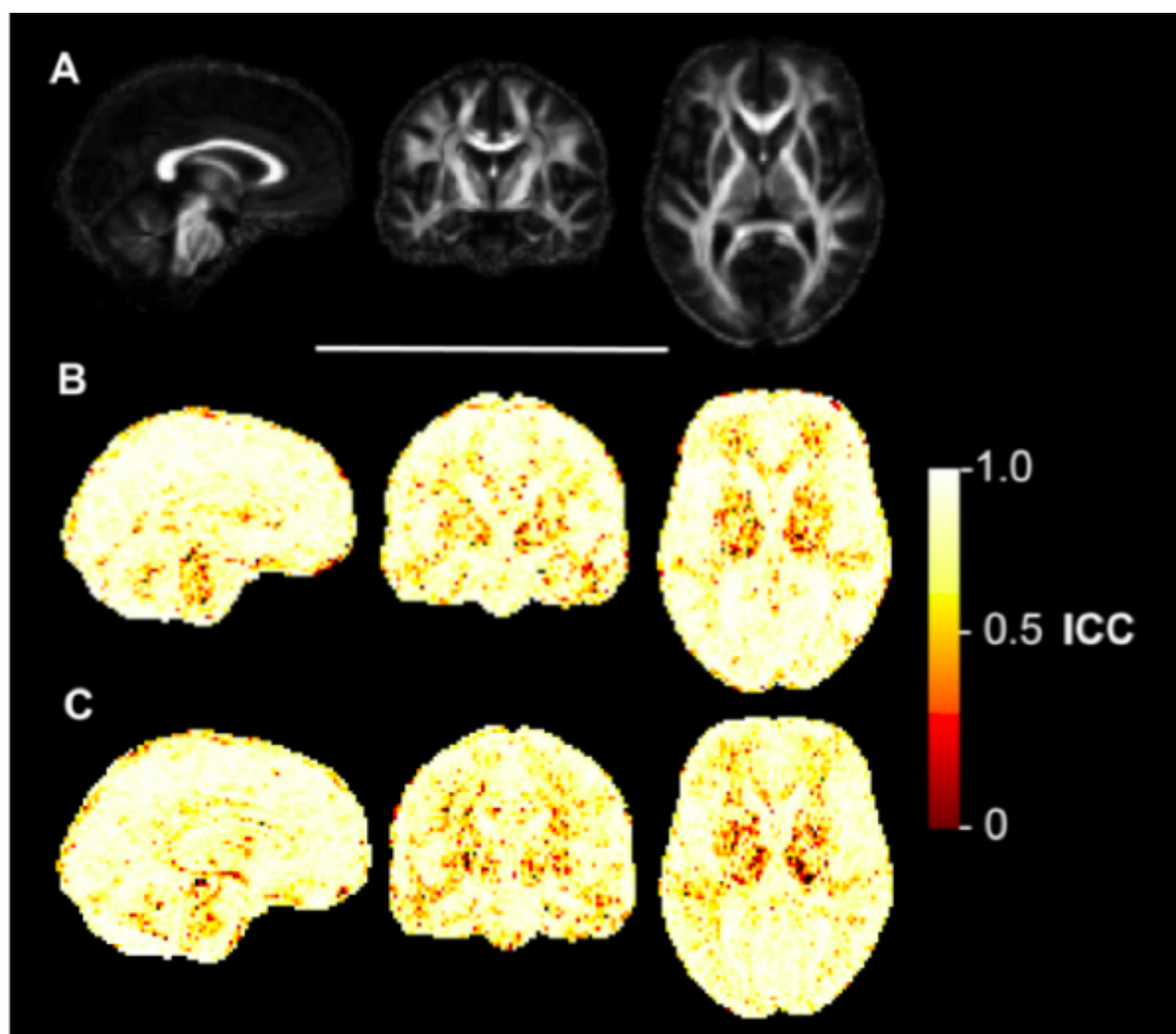




**Fig. 6: Voxelwise ICC maps for MD**

Voxelwise distribution of reliability metrics for mean diffusivity (MD). Panel A) shows sagittal, coronal and axial slices of the mean FA image created during the image processing, included for anatomical reference. B) Equivalent three slices for the intraclass correlation coefficient (ICC) of MD in early HD patients. Higher values reflect higher test-retest reliability. C) ICC of MD in healthy controls.

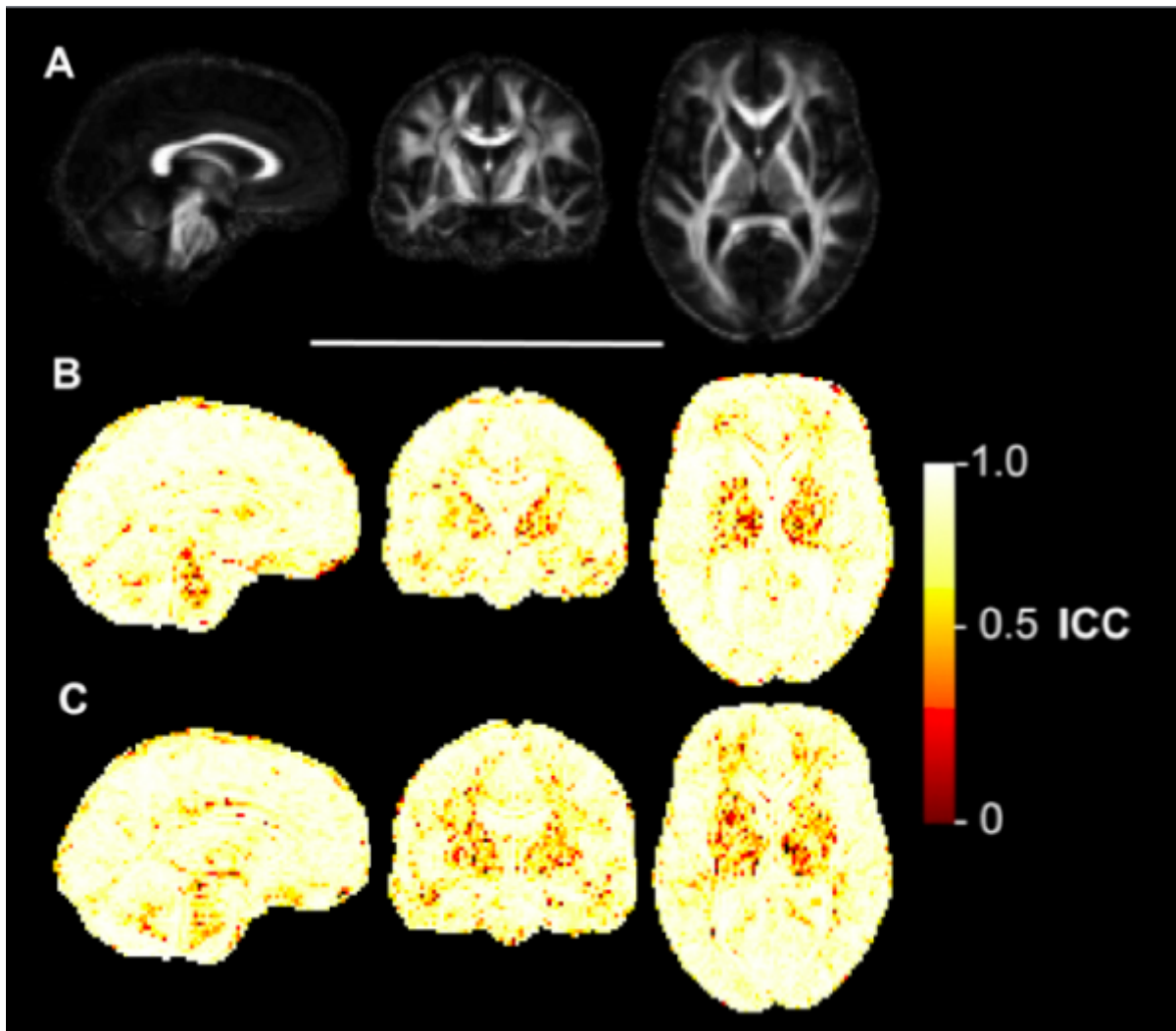




**Fig. 7: Voxelwise ICC maps of AD**

Voxelwise distribution of reliability metrics for axial diffusivity (AD). Panel A) shows sagittal, coronal and axial slices of the mean FA image created during the image processing, included for anatomical reference. B) Equivalent three slices for the intraclass correlation coefficient (ICC) of AD in early HD patients. Higher values reflect higher test-retest reliability. C) ICC of AD in healthy controls.





**Fig. 8: Voxelwise ICC maps of RD**

Voxelwise distribution of reliability metrics for radial diffusivity (RD). Panel A) shows sagittal, coronal and axial slices of the mean FA image created during the image processing, included for anatomical reference. B) Equivalent three slices for the intraclass correlation coefficient (ICC) of RD in early HD patients. Higher values reflect higher test-retest reliability. C) ICC of RD in healthy controls.

## Discussion

Analysis of DTI scans acquired back-to-back gave generally high levels of reliability in both early stage HD patients and controls, when using either atlas-based or T1-segmentation-based ROIs. Using specific methods that will be applied to longitudinal datasets in studies of HD, the four major DTI metrics (FA, MD, AD and RD) all showed low levels of within-subject (i.e. scan-rescan) variability relative to between-subject variability. This concurs with previous research that has assessed the reliability of DTI in HD<sup>43</sup> and other samples using different analysis methods, which generally report acceptable levels of reliability (i.e. high ICC), either test-retest<sup>24,35</sup> or between-scanners<sup>21,22,36,37</sup>. All metrics tended to perform as well in both experimental groups and importantly there was no qualitative evidence of group bias in reliability. Although no formal statistical comparisons were made between HD and controls, the ICC estimates and 95% CI suggested that for most regions the ICCs were consistent between groups. Given the magnitude of DTI effect sizes in cross-sectional comparisons of early HD patients and controls (e.g.<sup>4</sup>), the low within-subject variances found support the continued use of DTI as a tool for detecting patterns of neurodegeneration in this HD population.

It is important to recognise the imprecision in the ICCs reflected in the relatively wide confidence intervals, particularly for the lower ICCs. Nonetheless it does appear that levels of reliability were not entirely consistent across the brain. Those regions that showed lower reliability (ICC < 0.8) tended to be regions in the inferior of the brain, including cerebellum and brainstem ROIs. Explanations for this could include the increased presence of tissue susceptibility errors in inferior regions, or sub-optimal image registration due to the registration procedure being largely driven by the increased contrast in signal intensity found in and around the lateral ventricles<sup>32</sup>. Also, these cerebellar regions have smaller volumes when defined according to the ICBM-DTI-81 atlas, thus may be more susceptible to partial volume effects than larger areas where such effects are more likely to be averaged out. The present findings indicate that extra caution should be exercised when examining patterns of longitudinal change in small inferior regions, as these tend to have greater inherent variability using atlas-based registration methods.

Interestingly, the grey matter ROIs, segmented on a T1 image, showed lower reliability for diffusion metrics, particularly MD in the caudate and AD in both caudate and putamen. Regional atrophy may contribute to inaccuracies in the registration of T1 images to FA maps and the caudate in particular is susceptible to partial volume effects due to its proximity to the lateral ventricles. The process of transforming T1 ROIs to diffusion space is performed regularly in DTI ROI analyses and although not routinely reported, visual quality assurance of registration accuracy is absolutely necessary as achieving precise anatomical correspondence between images acquired using different modalities is challenging.

Although the low ICC values in the striatal ROIs may be some cause for concern for longitudinal studies, this may be explained by the presence of particularly low between-subject variability, meaning that minor divergences within a small number of participants can lead to large fluctuations of the estimated ICC. The influence of such outliers, particularly in light of the reduced control sample size for the T1-segmented regions (N = 8), could be pronounced. To assess this, outliers identified from scatter plots were removed and the ICCs recalculated. In most cases, the ICC increased on removal, though not necessarily to the high values observed for other regions. As an example, AD in the left corticospinal tract had a low ICC in the HD subjects. Removing a single outlier increased the ICC from 0.471 to 0.763, which is still lower than the value of 0.889 observed for the same region in the controls, or 0.964, found in HD patients in the right hemisphere.

The within-subject variance should also be considered when comparing between regions or metrics. While in general within-variance was strongly related to ICC, as would be expected, there are some regions, such as the fornix, where ICC was high but within-subject variance appeared much greater than other regions. The magnitude of the within-subject variance is approximately inversely proportional to the amount of signal, so for example, one may be cautious about the reliability of the fornix for longitudinal studies, despite a relatively high ICC of 0.90 for MD. If wishing to restrict the number of brain regions tested, areas such as the genu and body of corpus callosum, for which we observed high ICC and very low within-subject variance, might be more appropriate for longitudinal research studies.

Voxelwise ICC maps of FA, MD, AD and RD indicated that the distribution of variability was qualitatively very similar between early HD patients and controls. High ICC for FA was apparent across the white matter of the brain, with reduced values in the lateral ventricles and subcortical grey matter nuclei. The strength of the signal in DTI depends on the degree of anisotropy<sup>38</sup> and FA is a direct representation of this measure, so it is plausible that the inherent anisotropy of brain tissue influences the between-scan variability. The MD ICC maps also did not materially differ between groups, though when compared with the FA maps, the pattern was one of generally higher ICC throughout the brain, which was matched by very similar patterns for AD and RD; not unexpected given the relatedness of the metrics. The voxelwise patterns of reliability concur with the study by Marengo and colleagues<sup>24</sup> in healthy controls at 1.5T who showed a similar distribution of ICC for FA and trace (i.e. total diffusivity). In accordance with this, the voxelwise and ROI-based findings may add weight to the idea that DTI in general, and FA in particular, is primarily suited for examining white matter microstructure<sup>22</sup> and is less quantifiable<sup>39</sup> and harder to interpret in grey-matter regions<sup>40</sup>. Measures of anisotropy in tissue not thought to be characteristically anisotropic may not give particularly meaningful insight into biological or pathological processes.

Although the four DTI metrics were generally comparable across the board, there was a trend for lower reliability in AD, compared with FA, MD or RD. One possible explanation for this finding is that AD, unlike any of the other metrics, is derived from a single component of the diffusion tensor. For FA, MD and RD there is a degree of averaging across tensor elements (i.e. eigenvalues) that may increase the signal-to-noise ratio and reduce variability, whereas AD is derived solely from the primary eigenvalue. Previous reliability studies have not reported on AD before and this potential difference with other diffusion metrics should be investigated further, particularly if AD results are to be analysed longitudinally.

There are a few caveats to consider when interpreting the present findings. The sample size was relatively small once divided into early HD patients and controls, thus resulting in increased susceptibility to the influence of outliers and reduced precision. One issue in extrapolating these findings to future longitudinal studies is the scan-rescan data was collected back-to-back, with the participant remaining in the scanner. This meant that the position of the head within the magnetic field was very similar between scans, which is less often the case with longer intervals where re-positioning is required. When running longitudinal studies, monitoring the consistency of head positioning at each acquisition could help reduce the variability that can be caused by subtle differences in orientation and slice positioning<sup>19,41</sup>. Another limitation is that these results are scanner specific. The reliability of DTI may differ between scanner manufacturers<sup>37</sup>, or show higher noise/within-subject variability in older machines that require servicing. This point is particularly relevant to multi-centre studies, which are likely to be increasingly necessary in relatively uncommon diseases such as HD, in order to have well-powered studies. While the reliability of DTI has been demonstrated across scanners in principle<sup>36</sup>, collecting repeat scans both within-scanner and across study sites would be a prudent step to help establish the reliability of DTI data in any large scale study.

In conclusion, test-retest reliability analysis of atlas-based and T1-segmented ROI approaches DTI analysis show generally high ICCs on diffusion data acquired with a clinically-acceptable scan time. This gives confidence that such data acquisition and analysis methods can be reliably used for cross-sectional comparisons, alongside lending support for their utility to measure within-subject change over time; the goal of on-going longitudinal research into progressive neurodegeneration in HD. However, longitudinal reliability can only be explicitly demonstrated by taking the expected magnitude of the longitudinal effects into account, to determine whether these effects are greater than the scan-rescan variability. Finally, it is notable that there are some inconsistencies to the generally high reliability; particularly in striatal AD measures and it could be concluded that test-retest reliability is not evenly distributed throughout the brain, potentially due to intrinsic tissue differences, non-linear geometric distortions or uneven registration accuracy. This has implications for selecting which brain regions are most appropriate for future longitudinal studies, above and beyond the biological evidence for involvement in neurodegeneration.

## ***Competing Interests***

The authors have declared that no competing interests exist.

## Acknowledgements

In particular, the authors would like to acknowledge Prof. Sarah Tabrizi of UCL Institute of Neurology, as Principal Investigator of the PADDINGTON project and for her support of our work. The authors would also like to thank the patients and controls who took part in this study, along with all the Work Package 2 site staff at Paris, Leiden, Ulm and London.

## References

1. Bohanna I, Georgiou-Karistianis N, Sritharan A, Asadi H, Johnston L, Churchyard A, Egan G. Diffusion tensor imaging in Huntington's disease reveals distinct patterns of white matter degeneration associated with motor and cognitive deficits. *Brain Imaging Behav.* 2011 Sep;5(3):171-80. PubMed PMID:21437574.
2. Della Nave R, Ginestroni A, Tessa C, Giannelli M, Piacentini S, Filippi M, Mascalchi M. Regional distribution and clinical correlates of white matter structural damage in Huntington disease: a tract-based spatial statistics study. *AJNR Am J Neuroradiol.* 2010 Oct;31(9):1675-81. PubMed PMID:20488902.
3. Douaud G, Behrens TE, Poupon C, Cointepas Y, Jbabdi S, Gaura V, Golestani N, Krystkowiak P, Verny C, Damier P, Bachoud-Lévi AC, Hantraye P, Remy P. In vivo evidence for the selective subcortical degeneration in Huntington's disease. *Neuroimage.* 2009 Jul 15;46(4):958-66. PubMed PMID:19332141.
4. Hobbs NZ, Cole JH, Farmer RE, Rees EM, Crawford HE, Malone IB, Roos RA, Sprengelmeyer R, Durr A, Landwehrmeyer B, Scahill RI, Tabrizi SJ, Frost C. Evaluation of multi-modal, multi-site neuroimaging measures in Huntington's disease: Baseline results from the PADDINGTON study. *Neuroimage Clin.* 2012 Dec 9;2:204-11. PubMed PMID:24179770.
5. Delmaire C, Dumas EM, Sharman MA, van den Bogaard SJ, Valabregue R, Jauffret C, Justo D, Reilmann R, Stout JC, Craufurd D, Tabrizi SJ, Roos RA, Durr A, Lehericy S. The structural correlates of functional deficits in early huntington's disease. *Hum Brain Mapp.* 2013 Sep;34(9):2141-53. PubMed PMID:22438242.
6. Magnotta VA, Kim J, Kosciak T, Beglinger LJ, Espinso D, Langbehn D, Nopoulos P, Paulsen JS. Diffusion Tensor Imaging in Preclinical Huntington's Disease. *Brain Imaging Behav.* 2009 Mar 1;3(1):77-84. PubMed PMID:21415933.
7. Reading SA, Yassa MA, Bakker A, Dziorny AC, Gourley LM, Yallapragada V, Rosenblatt A, Margolis RL, Aylward EH, Brandt J, Mori S, van Zijl P, Bassett SS, Ross CA. Regional white matter change in pre-symptomatic Huntington's disease: a diffusion tensor imaging study. *Psychiatry Res.* 2005 Oct 30;140(1):55-62. PubMed PMID:16199141.
8. Rosas HD, Tuch DS, Hevelone ND, Zaleta AK, Vangel M, Hersch SM, Salat DH. Diffusion tensor imaging in presymptomatic and early Huntington's disease: Selective white matter pathology and its relationship to clinical measures. *Mov Disord.* 2006 Sep;21(9):1317-25. PubMed PMID:16755582.
9. Tabrizi SJ, Reilmann R, Roos RA, Durr A, Leavitt B, Owen G, Jones R, Johnson H, Craufurd D, Hicks SL, Kennard C, Landwehrmeyer B, Stout JC, Borowsky B, Scahill RI, Frost C, Langbehn DR. Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *Lancet Neurol.* 2012 Jan;11(1):42-53. PubMed PMID:22137354.
10. Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, Borowsky B, Landwehrmeyer B, Frost C, Johnson H, Craufurd D, Reilmann R, Stout JC, Langbehn DR. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurol.* 2013 Jul;12(7):637-49. PubMed PMID:23664844.
11. Sritharan A, Egan GF, Johnston L, Horne M, Bradshaw JL, Bohanna I, Asadi H, Cunningham R, Churchyard AJ, Chua P, Farrow M, Georgiou-Karistianis N. A longitudinal diffusion tensor imaging study in symptomatic



Huntington's disease. *J Neurol Neurosurg Psychiatry*. 2010 Mar;81(3):257-62. PubMed PMID:19237387.

12. Vandenberghe W, Demaerel P, Dom R, Maes F. Diffusion-weighted versus volumetric imaging of the striatum in early symptomatic Huntington disease. *J Neurol*. 2009 Jan;256(1):109-14. PubMed PMID:19267169.

13. Weaver KE, Richards TL, Liang O, Laurino MY, Samii A, Aylward EH. Longitudinal diffusion tensor imaging in Huntington's Disease. *Exp Neurol*. 2009 Apr;216(2):525-9. PubMed PMID:19320010.

14. Domínguez D JF, Egan GF, Gray MA, Poudel GR, Churchyard A, Chua P, Stout JC, Georgiou-Karistianis N. Multi-modal neuroimaging in premanifest and early Huntington's disease: 18 month longitudinal data from the IMAGE-HD study. *PLoS One*. 2013 Sep 16;8(9):e74131. PubMed PMID:24066104.

15. Rees, EM, Scahill RI, Hobbs NZ. Longitudinal Neuroimaging Biomarkers in Huntington's Disease. *Journal of Huntington's Disease*. 2013 2(1):21-39. DOI: 10.3233/JHD-120030.

#### [REFERENCE LINK](#)

16. Skare ST, Bammer R. EPI-Based Pulse Sequences for Diffusion Tensor MRI. In: Jones DK, editor. *Diffusion MRI : theory, methods, and applications*. New York ; Oxford: Oxford University Press; 2011.

17. Tijssen RH, Jansen JF, Backes WH. Assessing and minimizing the effects of noise and motion in clinical DTI at 3 T. *Hum Brain Mapp*. 2009 Aug;30(8):2641-55. PubMed PMID:19086023.

18. Bisdas S, Bohning DE, Besenski N, Nicholas JS, Rumboldt Z. Reproducibility, interrater agreement, and age-related changes of fractional anisotropy measures at 3T in healthy subjects: effect of the applied b-value. *AJNR Am J Neuroradiol*. 2008 Jun;29(6):1128-33. PubMed PMID:18372415.

19. Bonekamp D, Nagae LM, Degaonkar M, Matson M, Abdalla WM, Barker PB, Mori S, Horská A. Diffusion tensor imaging in children and adolescents: reproducibility, hemispheric, and age-related differences. *Neuroimage*. 2007 Jan 15;34(2):733-42. PubMed PMID:17092743.

20. Pfefferbaum A, Adalsteinsson E, Sullivan EV. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J Magn Reson Imaging*. 2003 Oct;18(4):427-33. PubMed PMID:14508779.

21. Teipel SJ, Reuter S, Stieltjes B, Acosta-Cabronero J, Ernemann U, Fellgiebel A, Filippi M, Frisoni G, Hentschel F, Jessen F, Klöppel S, Meindl T, Pouwels PJ, Hauenstein KH, Hampel H. Multicenter stability of diffusion tensor imaging measures: a European clinical and physical phantom study. *Psychiatry Res*. 2011 Dec 30;194(3):363-71. PubMed PMID:22078796.

22. Vollmar C, O'Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, Duncan JS, Richardson MP, Koepp MJ. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *Neuroimage*. 2010 Jul 15;51(4):1384-94. PubMed PMID:20338248.

23. Huang L, Wang X, Baliki MN, Wang L, Apkarian AV, Parrish TB. Reproducibility of structural, resting-state BOLD and DTI data between identical scanners. *PLoS One*. 2012;7(10):e47684. PubMed PMID:23133518.

24. Marengo S, Rawlings R, Rohde GK, Barnett AS, Honea RA, Pierpaoli C, Weinberger DR. Regional distribution of measurement error in diffusion tensor imaging. *Psychiatry Res*. 2006 Jun 30;147(1):69-78. PubMed PMID:16797169.

25. Brander A, Kataja A, Saastamoinen A, Ryymin P, Huhtala H, Ohman J, Soimakallio S, Dastidar P. Diffusion tensor imaging of the brain in a healthy adult population: Normative values and measurement reproducibility at 3 T and 1.5 T. *Acta Radiol*. 2010 Sep;51(7):800-7. PubMed PMID:20707664.

26. Shoulson I, Fahn S. Huntington disease: clinical care and evaluation. *Neurology*. 1979 Jan;29(1):1-3. PubMed PMID:154626.

27. Freeborough PA, Fox NC, Kitney RI. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Comput Methods Programs Biomed*. 1997 May;53(1):15-25. PubMed PMID:9113464.

28. Pierson R, Johnson H, Harris G, Keefe H, Paulsen JS, Andreasen NC, Magnotta VA. Fully automated analysis using BRAINS: AutoWorkup. *Neuroimage*. 2011 Jan 1;54(1):328-36. PubMed PMID:20600977.
29. Ourselin S, Roche A, Subsol G, Pennec X, Ayache N, Reconstructing a 3D structure from serial histological sections, *Image and Vision Computing*. 2001 Jan, 19(1):25-31.
30. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S. Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed*. 2010 Jun;98(3):278-84. PubMed PMID:19818524.
31. Wang Y, Gupta A, Liu Z, Zhang H, Escolar ML, Gilmore JH, Gouttard S, Fillard P, Maltbie E, Gerig G, Styner M. DTI registration in atlas based fiber analysis of infantile Krabbe disease. *Neuroimage*. 2011 Apr 15;55(4):1577-86. PubMed PMID:21256236.
32. Zhang H, Yushkevich PA, Alexander DC, Gee JC. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Med Image Anal*. 2006 Oct;10(5):764-85. PubMed PMID:16899392.
33. Mori S, Wakana S, Van Zijl PCM. MRI atlas of human white matter. 1st ed. Amsterdam, The Netherlands ; San Diego, CA: Elsevier; 2005.
34. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986 Feb 8;1(8476):307-10. PubMed PMID:2868172.
35. Jansen JF, Kooi ME, Kessels AG, Nicolay K, Backes WH. Reproducibility of quantitative cerebral T2 relaxometry, diffusion tensor imaging, and 1H magnetic resonance spectroscopy at 3.0 Tesla. *Invest Radiol*. 2007 Jun;42(6):327-37. PubMed PMID:17507802.
36. Magnotta VA, Matsui JT, Liu D, Johnson HJ, Long JD, Bolster BD Jr, Mueller BA, Lim K, Mori S, Helmer KG, Turner JA, Reading S, Lowe MJ, Aylward E, Flashman LA, Bonett G, Paulsen JS. Multicenter reliability of diffusion tensor imaging. *Brain Connect*. 2012;2(6):345-55. PubMed PMID:23075313.
37. Fox RJ, Sakaie K, Lee JC, Debbsins JP, Liu Y, Arnold DL, Melhem ER, Smith CH, Philips MD, Lowe M, Fisher E. A validation study of multicenter diffusion tensor imaging: reliability of fractional anisotropy and diffusivity values. *AJNR Am J Neuroradiol*. 2012 Apr;33(4):695-700. PubMed PMID:22173748.
38. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. *Radiology*. 1996 Dec;201(3):637-48. PubMed PMID:8939209.
39. Farrell JA, Landman BA, Jones CK, Smith SA, Prince JL, van Zijl PC, Mori S. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. *J Magn Reson Imaging*. 2007 Sep;26(3):756-67. PubMed PMID:17729339.
40. Rulseh AM, Keller J, Tintěra J, Kožíšek M, Vymazal J. Chasing shadows: what determines DTI metrics in gray matter regions? An in vitro and in vivo study. *J Magn Reson Imaging*. 2013 Nov;38(5):1103-10. PubMed PMID:23440865.
41. Virta A, Barnett A, Pierpaoli C. Visualizing and characterizing white matter fiber structure and architecture in the human pyramidal tract using diffusion tensor MRI. *Magn Reson Imaging*. 1999 Oct;17(8):1121-33. PubMed PMID:10499674.
42. Penney JB Jr, Vonsattel JP, MacDonald ME, Gusella JF, Myers RH. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann Neurol*. 1997 May;41(5):689-92. PubMed PMID:9153534.
43. Müller HP, Grön G, Sprengelmeyer R, Kassubek J, Ludolph AC, Hobbs N, Cole J, Roos RA, Duerr A, Tabrizi SJ, Landwehrmeyer GB, Süßmuth SD. Evaluating multicenter DTI data in Huntington's disease on site specific effects: An ex post facto approach. *Neuroimage Clin*. 2013;2:161-7. PubMed PMID:24179771.