SUPPLEMENTARY MATERIAL

SUPPLEMENTARY MATERIAL 1:

- **MRI sequences parameters**

Sequences		Repetition time (ms)	Echo time (ms)	Field of view (mm^2)	Voxel size (mm^3)	No. of excitations	No. of slices	Scan time (min)
Spinal Cord	Sagittal PD/T2-weighted turbo spin-echo	4000	15/80	256 x 160	1.0 x1.0 x 3.0	2	12	5:44
	Fat-suppressed 3D slab-selective fast field echo sequence	23	5	240 x 180	0.5 x 0.5 x 5	8	11	15:58
Brain	Axial PD/T2-weighted turbo spin-echo	3500	15/85	240 x 180	1 x 1 x 3	1	50	4:01
	Sagittal 3D T1-weighted magnetization-prepared gradient-echo sequence	6.9	3.1	256 x 256	1 x 1x 1	1	180	6:31

- **QSI acquisition protocol**

For the QSI, using the manufacturer's 32-channel head coil (Philips Medical Systems, Best, Netherlands), each subject underwent a cardiac gated DWI acquisition [parameters: voxel size = 1 x 1 x 5 mm3 (interpolated in k-space to a 0.5 x 0.5 mm2 in-plane resolution), FOV = 64 x 64 mm2; TR = 9RR, TE = 129ms]; 12 axial contiguous slices covering a 60 mm length of the cervical cord (from C1 to C3)]. A ZOOM (zonal oblique multi-slice) sequence with

outer-volume suppression was used to minimise artefacts (*Wilm BJ, Svensson J, Henning A, et al. Reduced field-of-view MRI using outer volume suppression for spinal cord diffusion imaging. Magn Reson Med 2007; 57: 625–630).* Thirty DWI volumes with equally spaced Q-values and two non-diffusion weighted (b0) volumes were acquired with diffusion weighting in two perpendicular (x and y) and one parallel (z) direction to the main axis of the spinal cord [parameters: diffusion pulse duration δ = 11.4ms, diffusion time Δ = 75ms, gradient strength G linearly increased in 31 steps from 0 to 87.5mT/m in x and y direction and 62mT/m in z direction]. We assumed axial symmetry of the axons along the long axis of the spinal cord, and therefore applied gradient amplifiers in two orthogonal directions, generating maximum gradient strengths of $\sqrt{2}$*62 mT/m and 62 mT/m in the xy and z direction, respectively. Q-values were the same in the xy and z directions, but smaller gradient pulse duration of 11.4 ms in the xy direction (16 ms in z) was used.

SUPPLEMENTARY MATERIAL 2:

- **Inter and intra-rater variability of the imaging analysis calculation**

The raters who placed the ROI in the brain and spinal cord were blinded to the subject identity and clinical characteristics but aware of the time points. To assess the reproducibility of the analysis and to ensure that results had not been influenced by different raters' analysis, a rater (R.C.), who was blinded to the original results, calculated CSA and repositioned the ROIs on the average b0 image in the cord for QSI analysis. This was done for 5 randomly selected patients at each time-point. To assess the inter and intra-rater variability of the imaging analysis, we obtained the Intraclass Correlation Coefficients (ICC) for each one of the measures. ICC is a widely used reliability index in intra-rater and inter-rater reliability analyses. Based on the 95% confident interval of the ICC estimate, values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively *(Koo & Li, 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med, 15(2), 155-163)*. In the analysis we performed, ICC was greater than 0.90 for all MRI measures we considered, indicating excellent reliability.

- **Intraclass Correlation Coefficients (ICC) for the MRI measures**

	BL		**1Y**		**3Y**	
	ICC	95% CI	ICC	95% CI	ICC	95% CI
CSA	0.9999	0.9998-0.9999	0.9996	0.9979-0.9999	0.9994	0.9964-0.9999
Cord b0	0.9989	0.9930-0.9998	0.9995	0.9972-0.9999	0.9999	0.9998-0.9999

*p-values for ICC calculation for all measures $p<0.05$