Since this is a very extensive supplemental material, the format and content have not been edited by JRM

Supplementary material

Appendix SI: Magnetic resonance imaging protocols

Brain MRI scans may induce geometric and intensity variabilities that adversely impact on automated measurements of lesion and brain volume change. To remedy this impact, we implemented the Magnetic Resonance Imaging in MS (MAGNIMS) guidelines (1). All brain MRI scans were acquired on the same 3T Philips Achieva scanner at the University of Ljubljana, Faculty of Medicine, Infrastructural centre of MR imaging and spectroscopy. The imaging protocol was the same for all study participants, and included: (*i*) T1-weighted images, (*ii*) T1-weighted images with gadolinium (Gd) enhancement, (*iii*) T2-weighted images, (*iv*) fluid-attenuated inversion recovery (FLAIR) images, and (v) dual inversion recovery (DIR) images. Vendor-provided geometric distortion correction using phantom-based calibration was applied. Furthermore, at the expense of extended acquisition time, averaging with a factor of 2 was employed to each scan so as to boost image quality. All sequences were high-resolution 3D, acquired in the sagittal direction. The sequences and corresponding parameters are listed in **Table SI**.

Table SI. List of acquired magnetic resonance (MR) sequences and corresponding parameters: echo time (TE), repetition time (TR), inversion time (TI) and flip angle (FA). All sequences were acquired in the sagittal direction

SEQUENCE	**SAMPLING** [pix]	**SPACING** [mm]	**TE** [ms]	**TR** [ms]	**TI** [ms]	**FA** [°]
T1	352×165×352	0.66×1.00×0.66	4.3	9.2	-	8
T1+Gd	352×165×352	0.66×1.00×0.66	4.3	9.2	-	8
T2	352×165×352	0.66×1.00×0.66	332	2500	-	90
FLAIR	240×321×240	0.97×0.64×0.97	276	4800	1650	90
DIR	256×256×300	0.97×0.64×0.97	249	5500	2550	90

T1: T1-weighted sequence; T2: T2-weighted sequence; FLAIR: Fluid Attenuated Inversion Recovery; DIR: Double Inversion Recovery; Gd: gadolinium.

Appendix SII: Image analysis

All MRI scans were quantified using MS Markers (UL FE, Tržaška 25, 1000 Ljubljana, Slovenia), an online automated image analysis software service (2). MS Markers extracted the contours of several brain structures and active, T2 and cortical lesions.

In MS Markers, all MRI scans were first pre-processed by applying non-local-means based image de-noising (3) and the N4 bias correction (4). From the pre-processed T1-weighted (T1) image the brain mask was extracted by a multi-atlas label fusion segmentation method (5), which employed 50 manually segmented T1 MRI brain images of age-matched healthy subjects. The atlases were aligned to the brain masked and pre-processed T1 by a non-linear B-spline registration method (6). Using the same registration method, the corresponding pre-processed Gd-enhanced T1, FLAIR, and DIR images were aligned to the pre-processed T1. Also, the T1 image from an MNI 2009c non-linear symmetrical brain atlas (7) was aligned to the pre-processed T1-weighted image so as to determine a volumetric normalization factor.

On baseline MRI scans, an unsupervised method for brain tissue and hyper-intense white-matter (WM) or T2 lesion segmentation (8) was applied to the co-registered pairs of brain masked and pre-processed T1 and FLAIR images. Lesion-filling algorithm (9) was used to reconstruct normal-appearing tissue intensity on the pre-processed T1 image, and the tissue segmentation was re-run to obtain the final WM and GM segmentations. Using the multi-atlas label fusion segmentation method (5) and Neuromorphometrics brain atlases of 30 subjects (10) the following brain substructures were segmented in the lesion-filled T1 image: (*i*) lateral ventricles, (*ii*) thalamus, (*iii*) precentral gyrus, (*iv*) basal ganglia including the nucleus accumbens, pallidum, putamen and caudate nucleus, (*v*) the limbic lobe including the anterior, middle and posterior cingulate gyrus, and (*vi*) medial temporal lobe including amygdala, hippocampus and parahippocampus.

The T2 lesion segmentations were obtained along with the tissue segmentation. Cortical lesions were segmented from pairs of co-registered and pre-processed T1 and DIR images that were input into a modified unsupervised segmentation method (8), as used for the WM lesion segmentation. Namely, the hyper-intense DIR outliers, which contain the cortical lesions, were constrained to the GM region. The co-registered, brain masked and pre-processed T1 and Gd-enhanced T1 images were intensity co-normalized using differential bias correction method (11) and subtracted. From the subtraction images the Gd-enhancing lesions were segmented based on the intensity z-score threshold exceeding value of 3.0, and constrained within the WM region.

The pre-processed T1, Gd-enhanced T1, FLAIR and DIR follow-up MRI scans were aligned to corresponding baseline scans using affine registration method. To assess atrophy changes for structures segmented from pairs of baseline and follow-up T1 images, the B-spline registration was also performed in a 2-step sequence using 4- and 2-mm control grid spacing. Atrophy was computed from Jacobian determinant of the obtained B-spline deformation field by integrating the values across the segmentation of each structure of interest (12). To assess changes in Gd-enhancing, T2, and cortical lesions from Gd-enhanced T1, FLAIR and DIR images, respectively, the baseline and follow-up scans were intensity co-normalized (11) and subtracted. The lesion change segmentations were based on the intensity z-score threshold exceeding absolute value of 3.0, and constrained to the WM in case of Gd-enhancing and T2 lesions, and to GM in case of cortical lesions (13).

Before extracting the brain structures' and lesion volumes and lesion counts all the segmentations obtained from MS Markers were revised by the neuroradiologist in a blinded manner and, if required, manually updated using interactive visualization and segmentation tools and protocols, as described in Lesjak et al. (14). The baseline and baseline-to-follow-up MR measurements were then extracted from the revised segmentations. By multiplying by the volumetric normalization factor and voxel sizes, the corresponding normalized volumes for the brain structures were computed. From the lesion and lesion change segmentations the corresponding lesion volumes and counts according to 18-connected object labelling were computed.

Appendix SIII: Cognitive testing

Verbal learning and delayed memory were tested by CVLT-II. After each verbal presentation of the word list, participants had to repeat as many words as possible. The total number of words recalled after the fifth word list presentation was measured.

The SDMT required participants to mentally manipulate symbols and numbers in a limited time interval. The SDMT provides a measure of complex attention, combined with psychomotor speed, visuomotor coordination and working memory. A higher number of correct associations between numbers and symbols indicates an effective complex attention. The measured variable was the total number of correct responses in 90 s.

Finally, BVMT-R was used to measure visuospatial learning and delayed memory. After each presentation of some simple visuospatial designs, participants had to reproduce as many of the designs as possible. The total number of scores after 3 consecutive representations was recorded.

To avoid potential confound and ensure test-retest reliability, different versions of CVLT-II and BVMT-R were used on each occasion, while there are no known contradictions or limits to reusing the SDMT again after a 12-week time interval.

Appendix SIV: MSQOL-54 subscore results

Table SII. Estimated mean MSQoL-54 subscale outcomes

Outcome measures	Group	Baseline-adjusted means Mean (95% CI)	Adjusted means at 3-month follow-up Mean (95% CI)	Mean differences between groups at 3-month follow-up	
				Mean (95% CI)	*p*-value
Physical health	Intervention	72.42 (35.87, 108.98)	84.95 (49.53, 120.37)	–2.69 (–10.49, 5.10)	0.506
	Control	70.79 (55.56, 86.03)	87.64 (52.09, 123.19)		
Health perceptions	Intervention	56.02 (35.81, 76.23)	53.07 (20.17, 85.97)	–0.28 (–11.07, 10.51)	0.960
	Control	54.90 (35.42, 74.39)	53.35 (39.36, 67.35)		
Energy	Intervention	56.19 (23.77, 88.62)	47.27 (19.66, 74.89)	–10.95 (–20.84, –1.07)	0.040
	Control	49.52 (36.01, 63.04)	58.23 (46.31, 70.14)		
Role limitations due to physical problems	Intervention	61.64 (10.10, 113.17)	61.37 (–27.62, 150.36)	–6.04 (–39.32, 27.24)	0.725
	Control	54.09 (10.15, 98.02)	67.41 (21.96, 112.85)		
Pain	Intervention	67.62 (22.24, 113.01)	80.07 (44.60, 115.55)	3.36 (–8.27, 14.98)	0.577
	Control	69.88 (50.96, 88.79)	76.72 (61.63, 91.80)		
Sexual function	Intervention	78.20 (16.88, 139.51)	85.88 (71.19, 100.57)	1.27 (–7.88, 10.43)	0.788
	Control	77.78 (52.22, 103.33)	84.61 (72.06, 97.15)		
Social function	Intervention	79.63 (61.84, 97.42)	76.80 (57.84, 95.76)	–9.56 (–18.81, –0.32)	0.054
	Control	76.63 (60.69, 92.57)	86.37 (73.05, 99.68)		
Health distress	Intervention	76.76 (52.39, 101.12)	84.57 (70.76, 98.39)	1.50 (–7.17, 10.17)	0.738
	Control	81.67 (65.82, 97.53)	83.08 (71.02, 95.13)		
Overall quality of life	Intervention	70.45 (28.44, 112.46)	64.99 (29.03, 100.94)	–7.80 (–19.26, 3.65)	0.194
	Control	71.03 (53.52, 88.54)	72.79 (58.08, 87.50)		
Cognitive function	Intervention	86.34 (47.57, 125.12)	81.97 (71.53, 92.42)	3.31 (–2.77, 9.40)	0.296
	Control	77.58 (61.42, 93.75)	78.66 (71.08, 86.24)		

Emotional well-being	Intervention Control	77.68 (53.18, 102.17) 75.37 (60.92, 89.82)	66.66 (46.41, 86.92) 73.36 (60.19, 86.52)	–6.70 (–16.73, 3.34)	0.203
Role limitations due to emotional problems	Intervention Control	70.84 (10.85, 130.83) 95.80 (65.17, 126.43)	82.70 (8.60, 156.81) 72.06 (42.23, 101.88)	10.65 (–14.20, 35.49)	0.409
Satisfaction with sexual function	Intervention Control	58.40 (–9.61, 126.42) 73.74 (45.39, 102.09)	73.45 (33.03, 113.87) 69.41 (48.33, 90.48)	4.05 (–11.95, 20.05)	0.625
Change in health	Intervention Control	47.73 (–11.04, 106.51) 61.20 (36.70, 85.70)	49.17 (1.06, 97.27) 54.41 (33.89, 74.92)	–5.24 (–22.59, 12.11)	0.559

95% CI: 95% confidence interval.

SUPPLEMENTARY REFERENCES

1. Rovira À, Wattjes MP, Tintoré M, et al. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-clinical implementation in the diagnostic process. Nat Rev Neurol. 2015 Aug; 11: 483. Corrected and republished from: Nat Rev Neurol 2015; 11: 471–482.
2. MS Markers. UL FE, Tržaška 25, 1000 Ljubljana, 2017. Accessed on: 9.3.2021. Available from: https://ms.quantim.eu/
3. Manjón JV, Coupé P, Buades A, Louis Collins D, Robles M. New methods for MRI denoising based on sparseness and self-similarity. Med Image Anal 2012; 16: 18–27.
4. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010; 29: 1310–1320.
5. Cardoso MJ, Modat M, Wolz R, Melbourne A, Cash D, Rueckert D, et al. Geodesic information flows: spatially-variant graphs and their application to segmentation and fusion. IEEE Trans Med Imaging 2015; 34: 1976–1988.
6. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Trans Med Imaging. 2010; 29:196–205.
7. Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins DL. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage 2009; 47: Suppl: S102.
8. Galimzianova A, Pernuš F, Likar B, Špiclin Ž. Stratified mixture modeling for segmentation of white-matter lesions in brain MR images. Neuroimage 2016; 124 (Pt A): 1031–1043.
9. Chard DT, Jackson JS, Miller DH, Wheeler-Kingshott CAM. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 2010; 32: 223–228.
10. MICCAI 2012 Grand challenge and workshop on multi-atlas labeling. 2017. Accessed on: 9.3.2021. Available from: http://www.neuromorphometrics.com/2012_MICCAI_Challenge_Data.html
11. Lewis EB, Fox NC. Correction of differential intensity inhomogeneity in longitudinal MR images. Neuroimage 2004; 23: 75–83.
12. Pai A, Sporring J, Darkner S, Dam EB, Lillholm M, Jørgensen D, et al. Deformation-based atrophy computation by surface propagation and its application to Alzheimer's disease. J Med Imaging 2016; 3: 014005.
13. Lesjak Ž, Pernuš F, Likar B, Špiclin Ž. Validation of white-matter lesion change detection methods on a novel publicly available MRI image database. Neuroinformatics 2016; 14: 403–20.
14. Lesjak Ž, Galimzianova A, Koren A, Lukin M, Pernuš F, Likar B, et al. A novel public MR image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. Neuroinformatics 2018;16(1):51-63.