Title: Networks of microstructural damage predict disability in multiple sclerosis

Elisa Colato, PhD, [1] Ferran Prados, PhD,[1,7,8,9] Jonathan Stutters[1], Alessia Bianchi, MD,[1], Sridar Narayanan, PhD, [3] Douglas L. Arnold, MD,[3] Claudia Gandini A.M. Wheeler-Kingshott, PhD,[1,4,5] Frederik Barkhof, MD, PhD,[1,6,7,9] Olga Ciccarelli, PhD, FRCP,[1,7] Declan T. Chard, PhD, FRCP[1,7] and Arman Eshaghi, MD, PhD[1,2]

Affiliations:

1 Queen Square Multiple Sclerosis Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, UK

2 Centre for Medical Image Computing (CMIC), Department of Computer Science, University College London, UK

3 McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

4 Brain Connectivity Centre, IRCCS Mondino Foundation, Pavia, Italy

5 Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

6 Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, location Vrije Universiteit, Amsterdam, NL

7 Institute for Health Research (NIHR), University College London Hospitals (UCLH) Biomedical Research Centre (BRC), London, UK

8 e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain

9 Centre for Medical Image Computing (CMIC), Department of Medical Physics and Biomedical Engineering, University College London, UK

Correspondence to:

Elisa Colato, PhD (corresponding author on behalf of co-authors)

Queen Square Multiple Sclerosis, UCL Queen Square Institute of Neurology

University College London, London, WC1B5EH, United Kingdom

Telephone: 0044 203 108 7420

elisa.colato.18@ucl.ac.uk

Running title: Predicting disability with brain networks in MS

Keywords: disability progression; multiple sclerosis, independent component analysis, cognitive impairment, brain network

Abbreviations:

MS = Multiple Sclerosis; 24-week CDP = Confirmed Disability Progression at 24 weeks; AIC = Akaike's Information Criterion (AIC); C-index = Concordance Index; CI = Confidence Interval; EDSS = Expanded Disability Status Scale; GM = Grey Matter; HR = Hazard Ratio; ICA = Independent Component Analysis; MRI = Magnetic Resonance Imaging; NAGM = Normal-Appearing Grey Matter; NAWM = Normal-appearing White Matter; PPMS = Primary Progressive Multiple Sclerosis; RRMS = Relapsing-Remitting Multiple Sclerosis; SDMT = Symbol Digit Modalities Test; SPMS = Secondary Progressive Multiple Sclerosis; sT1w/T2w = Standardised T1-weighted/T2-weighted ratio; WM = White Matter

Word count: 3975

Reference count: 47

Abstract

Background. Network-based measures are emerging MRI markers in multiple sclerosis (MS). We aimed to identify networks of white (WM) and grey matter (GM) damage that predict disability progression and cognitive worsening using data-driven methods.

Methods. We analysed data from 1836 participants with different MS phenotypes (843 in a discovery cohort, and 842 in a replication cohort). We calculated standardised T1/T2 (sT1w/T2w) ratio maps in brain GM and WM, and applied spatial independent component analysis to identify networks of covarying microstructural damage. Clinical outcomes were Expanded Disability Status Scale (EDSS) worsening confirmed at 24 weeks (24-week CDP) and time to cognitive worsening assessed by the Symbol Digit Modalities Test (SDMT). We used Cox proportional hazard models to calculate predictive value of network measures.

Results. We identified 8 WM and 7 GM sT1w/T2w networks (of regional co-variation in sT1w/T2w measures) in both cohorts. Network loading represents the degree of co-variation in regional T1/T2 ratio within a given network. The loading factor in the anterior corona radiata and temporo-parieto-frontal components were associated with higher risks of developing CDP both in the discovery (respectively, HR=0.85, P<0.05; and HR=0.83, P<0.05) and replication cohorts (respectively, HR=0.84, P<0.05; and HR=0.80, P<0.005). The decreasing or increasing loading factor in the arcuate fasciculus, corpus callosum, deep GM (DGM), cortico-cerebellar patterns, and lesion load were associated with a higher risk of developing SDMT worsening both in the discovery (HR=0.82, P<0.01; HR=0.87, P<0.05; HR=0.75, P<0.001; HR=0.86, P<0.05; and HR=1.27, P<0.0001) and replication cohorts (HR=0.82, P<0.005; HR=0.73, P<0.0001; HR=0.80, P<0.005; HR= 0.85, P<0.01; and HR= 1.26, P<0.0001).

Conclusions. GM and WM networks of microstructural changes predict disability and cognitive worsening in MS. Our approach may be used to identify patients at greater risk of disability worsening and stratify cohorts in treatment trials.

Highlights

- Network measures complement regional and global white and grey matter measures in explaining disability and predicting its worsening.

- This study is the first study to use networks of changes in standardised T1w/T2w, a proxy for microstructural damage, in WM and GM to predict the disability in MS.

- We found clinically relevant GM and WM networks of microstructural changes, some of which predicted the disability progression and cognitive worsening in a large cohort of participants with MS.

- The identification of networks of GM and WM changes, based on MRI scans routinely collected in clinical trials, that predict progression may be used in combination with other factors to identify treatment trial participants most likely to experience disability progression.

- Because sT1w/T2w maps were estimated from MRI scans routinely acquired in clinical trials, our models have the potential to be widely applied to future clinical trials, identifying participants at higher risk of progression.

Introduction

Multiple sclerosis (MS) is an inflammatory, demyelinating, and neurodegenerative disease. Most people with MS accumulate irreversible neurological and cognitive disability.[1] Markers that predict clinical and cognitive progression are needed in MS[1] as they can have useful roles in clinical trials and in clinical practice.

Previous studies quantifying the risk of disability worsening using MRI have mostly focused on white matter (WM) lesions (number and location) and relapses, especially in the relapsing-remitting MS.[2,3] Neurodegeneration, manifest in part as brain grey matter (GM) atrophy,[4] is also recognised as a leading cause of long-term irreversible disability, particularly in progressive MS.[5] However, it is now well-recognised that WM lesions and brain atrophy represent only a fraction of the MS pathology, and that they do not fully explain clinical disability in MS.[6] Pathology beyond lesions, in normal-appearing (NA) WM and in GM, goes undetected but can be assessed using advanced MRI [7–9] , and this may help further bridge the gap between clinical and radiological findings.

Advanced MRI sequences are not routinely obtained in clinical practice or in large clinical trials. T1/T2 ratio measures are a promising marker of microstructural damage[10,11] because they can be estimated from MRI sequences routinely acquired in clinical practice and trials.[12] The pathological substrates of the T1/T2 ratio are still debated and it is not yet clear whether this measure reflects demyelination, axonal or dendrites loss.[13–18] Previous studies have found a correlation between T1/T2 ratio maps and histological cortical myeloarchitectural. [17] T1/T2 ratio measures correlate with magnetization transfer ratio (MTR) measures (considered relatively specific for myelin content) in normal appearing WM (NAWM) and normal-appearing GM (NAGM), and T1/T2 has similar accuracy and sensitivity compared to MTR in detecting cortical demyelination.[12,19] In patients with a CIS, T1w/T2w alterations precede lesion formation and are associated with disease activity.[20] Recently, lower lesional and cortical T1w/T2w ratio values have been shown to be associated with longer disease duration, higher EDSS, higher brain lesional volume and lower normalised brain volume at all MS disease stages.[9,20]

Recent work using independent component analysis (ICA) of volumetric brain MRI scans has shown spatial patterns of cortical GM atrophy. This is likely to represent underlying network-

based pathology, an interpretation supported by patterns that resemble structural and functional networks.[21] These patterns of cortical changes may be related to the distribution of WM lesions[22] and with clinical outcomes in MS.[21,23–25] ICA is a data-driven (i.e. no a-priori hypothesis) algorithm that identifies spatially-independent patterns while looking for non-Gaussian spatial sources. It is based on the maximization of non Gaussianity, thus it forces the data to be as far from the normal distribution as possible, to identify patterns independent from each other. ICA offers the advantage over conventional atlas- or voxel- based analysis of investigating the inter-relationship between covarying voxels and provides a more holistic understanding of brain changes. Structural network-based analyses may better explain clinical outcomes and to complement conventional MRI measures.[25,26] We have recently shown that network-based measures of covarying GM volumes are associated with neurological disability measures and cognitive performance.[27] Similarly, ICA applied to diffusion tensor imaging (DTI) data has shown non-random patterns of fractional anisotropy changes in WM tracts associated with cognitive functioning in SPMS.[28] We hypothesise that data-driven networks based on T1/T2 can identify non-random covariation of WM and GM microstructural damage, and predict clinical and cognitive progression in MS.

Here we aimed to (1) identify networks of WM and GM in MS using standardised T1w/T2w ratio maps from MRI sequences acquired in clinical trials, (2) determine whether structural networks could be replicated in separate cohorts, (3) determine the clinical relevance of these measures by predicting EDSS and cognitive dysfunction, and (4) if the clinical relevance of these networks differed across clinical MS phenotypes.

Materials and methods

Overview

This was a retrospective study using cross-sectional MRI and longitudinal clinical data acquired at study entry. We used data from eight randomised-controlled clinical trials[29–35] acquired under the auspices of the International Progressive Multiple Sclerosis Alliance (IPMSA) and as part of the MS-SMART trial.

Participants

We included cross-sectional MRI, and cross-sectional and longitudinal clinical data from eight, randomized-controlled clinical trials. We balanced the data according to the clinical phenotype and clinical trial, and randomly sampled 1900 participants (RRMS, SPMS, and PPMS) from a database of 5089 participants to a discovery and replication cohorts to avoid bias. Computational resource limitations meant that we could not process data from the full cohort, and so we sub-selected 1900, which is larger than previous studies using sT1/T2 data effects.[7] We did not include all participants from all clinical trials because of memory requirements for the computational analysis. The IPMSA database was collected to target progressive MS trials, and 81% of patients in this study had progressive MS. Of the eight clinical trials sampled, three showed treatment effects,[30,32,36] for which we included data from the placebo arm only.

The Institutional Review Board at the Montreal Neurological Institute (MNI), Quebec, Canada approved this study (Reference number: IRB00010120). Participants gave informed consent to collect their data. All visits that fulfilled availability criteria of MRI and clinical data (explained below) were included.

To determine whether similar networks were present in healthy controls (HC), we performed additional analysis as described in supplemental materials.

Clinical and cognitive outcomes

We used the Expanded Disability Status Scale (EDSS) and Symbol Digit Modalities Test (SDMT) scores to assess disability and information processing speed. We estimated the EDSS worsening as an increase of 1 point from a baseline EDSS score of 5.5 or below, or of 0.5 points from baseline EDSS score greater than 5.5, and whether the worsening was confirmed

at 24 weeks (24-week confirmed disability progression or CDP).[37] We excluded clinical visits within 30 days of an MS relapse, where clinical attack dates were available. We defined cognitive worsening as a 4 point reduction in SDMT score.[38]

Brain MRI acquisition and image processing

Inclusion criteria for MRI were the availability of 2D or 3D, 1.5T or 3T, T1-weighted (T1w) without Gadolinium administration, T2-weighted (T2w), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) MRI scans. Details of MRI protocols are given in the supplemental materials and in the source clinical trial publications.[29,30,32–36]

Standardised T1w/T2w ratio maps (sT1w/T2w)

The first image processing goal was to obtain standardised T1w/T2w ratio maps. These scans were then used as input for the pattern analysis with ICA. We processed all available data for this study using the pipeline shown in Figure 1.

Standardised T1w/T2w ratio (sT1w/T2w) is a measure designed to address the technical and methodological limitations of the T1w/T2w ratio. T1/T2 ratio measures are derived from T1w and T2w intensities, which are variable and influenced by several technical and methodological factors (e.g. field strength or scanner manufacturer). The normalisation in sT1w/T2w measures addresses inhomogeneity in image intensities and harmonises measures for cross-subject comparisons.[7]

To obtain sT1w/T2w ratio maps in a common standard space, we followed an analysis pipeline similar to Cooper *et al.*[7] We also calculated sT1w/T2w ratios using another technique[9,39] on a randomly selected subsample to assess the consistency of measures between techniques and present the findings of this in the supplementary materials. Briefly, we resampled MRI images to 1x1x1, used N4 bias field correction tool in Advanced Normalization Toolbox (ANTs) to correct for bias field inhomogeneities in T1w and T2w scans.[40] We used NiftyReg to rigid register T1w and T2w images to a halfway space between the two modalities to avoid any bias toward one of the two modalities. We performed lesion segmentation using the DeepMedic software[41] on T2-FLAIR scans. We transformed the lesion masks and segmentation maps obtained with Geodesic Information Flows (GIF) version 3.0[42] to the mid-space, and subtracted lesion masks from WM masks to extract NAWM maps.

We applied an established formula to obtain sT1w/T2w.[7] Specifically, we estimated a scaling factor by dividing median values in GM in T1w by the median GM intensity in T2w. We obtained a scaled T2 image by multiplying the T2w scan by the scaling factor and estimated standardised T1w/T2w maps in the native space.

We transformed sT1w/T2w maps to a common space (customised study-specific template obtained as described in our previous work[27]) to allow the ICA to consistently identify covarying patterns. For further details on image processing, see Supplementary Materials.

We visually inspected the output of each step to check for segmentation errors (e.g. WM/GM maps estimation), and misregistration.

Network-based measures

We utilised the FastICA algorithm from scikit-learn 0.23.1[43] to identify spatial patterns of covarying 1) WM changes and 2) GM changes from sT1w/T2w ratio maps in the study-specific template. We refer to these spatial brain maps as networks in this manuscript. We specified the maximum number of components the ICA should attempt to extract at 20. This allowed subtle patterns to be extracted, albeit with caveat that these may produce measures which are noisier than the dominant networks of changes, and so have less potential to correlate with clinical outcomes. [44] We repeated the same for GM sT1w/T2w maps. To determine the stability of the independent components, we repeated the analysis independently in the replication cohort. We generated a 4D image by concatenating the 20 identified components and assessed pairwise spatial cross-correlations between ICs from the discovery and replication cohorts to select components that were spatially stable for each cohort. Higher correlation coefficients imply more spatial overlap between components across our two data sets. We considered ICs stable if there were statistically significant spatial voxel-wise correlations ($P<0.05$) across two data sets. To be even more conservative, we visually checked the identified components and exclude those not resembling well known functional or structural systems. We obtained the loading factor (i.e. the contribution of a subject to an independent component) for each participant in each network for statistical analysis for the discovery and replication cohort.

Because ICA loading factors will not always follow the same direction as variables underlying them (i.e. for a component both a positive or negative loading can represent decreasing or increasing sT1w/T2w), we performed Pearson correlations between the ICA loading for each network and sT1w/T2w measures. To aid readability, where needed we inverted the sign of the

ICA loading for a network so that for the final results a positive loading corresponded with increasing sT1w/T2w.

Statistical analysis

We used R (version 3.6.1) for statistical analysis. We assessed the demographic characteristic for each dataset and looked for differences across datasets and phenotypes (demographics and clinical) using the Mann-Whitney U test and two-sample t-test for continuous variables, and the Chi-Square test for categorical ones. We used two sample t-test to assess differences in WM and GM network measures between clinical phenotypes. We performed Pearson correlations to determine the relationship between WM and GM network measures with whole brain volume and lesion load.

Network measures

In the discovery cohort, we calculated the extent to which each participant contributed to each ICA component (i.e. "loading factor"). We calculated *z*-scores from the loading factor for each WM and GM-IC. We built a Cox regression model for each independent variable to determine whether baseline WM and GM networks, lesion load, and the volume of whole brain GM, thalamus, caudate, and pallidum (and the combination of the latter variables as "DGM") could predict EDSS and cognitive worsening. In these models, the event and the time-to-event were dependent variables. We adjusted each model for age, gender, clinical trial, MRI protocol, treatment, and disease duration. We repeated the analysis for each MS phenotype. We repeated the same analysis in the replication cohort. We corrected for multiple comparisons using Benjamini-Hochberg.

We performed post-hoc multivariate stepwise backward regression analysis including conventional MRI measures (i.e. lesion load and whole brain GM volume) and variables that were consistently identified from univariate Cox regression models to be associated with clinical and cognitive progression both in the discovery and replication cohorts.

Data and code availability

Data presented in this manuscript are controlled by various pharmaceutical companies. Therefore, data cannot be shared by investigators but can be requested directly from the pharmaceutical companies sponsoring each clinical trial. The computer code to obtain standardized T1w/T2w ratio maps in native space is available at https://github.com/co-el/Estimate-standardized-T1-T2-maps.

Results

Participants

MRI data were available for 1836 participants. We excluded 64 participants from the original sample: 31 had artefacts on the available images and T2w scans were missing for 33 participants. Clinical data were available for 1685 participants. The discovery cohort consisted of 843 participants with MS (310 men and 533 women with mean age of 46.6±9.3, mean disease duration of 10.2±8.4 years, mean follow-up of 2.75±1.74; 19% RRMS, 64%SPMS, and 17% PPMS). The replication cohort consisted of 842 participants (329 men and 513 women with mean age of 46.2±9.6, mean disease duration of 9.7±8.1 years, mean follow-up of 2.73±1.84; 19% RRMS, 64% SPMS, and 17% PPMS) with MS. Demographics are reported in Table 1 and in Supplementary Materials Table s1**.**

The discovery and replication cohorts were not different in age (P=0.39) and gender (P=0.36), lesion load (P=0.68), baseline EDSS (P=0.54), baseline SDMT (P=0.17), whole-brain GM volume (P=0.56), and disease duration (P=0.20). Comparisons across these measures remained non-significant between the discovery and replication cohorts when we looked individual MS phenotypes (Table 1).

Table 1. Demographics

Discovery Cohort				
Phenotype	ALL (n= 843)	RRMS (n= 159)	SPMS (n= 537)	PPMS (n= 147)
Gender (M/F)	310/533	48/111	178/359	84/63
Age (years)	46.6 ± 9.3	40.7±10.2	48.5±8.4	46.4±8.7
EDSS (median, range)	5.5 [0-8]	2.5 [0-8]	6 [3-7]	5 [2-6.5]
SDMT score (mean, SD)	42.3±14.7	48.0±18.8	41.5±13.9	NA
Disease duration (years)	10.2±8.4	7.1±8.9	12.9±7.9	3.4±3.6
Lesion load (ml)	28.5.0±25.0	23.8.3±24.7	31.2±25.5	23.7±22.1
Whole-brain GM volume (ml)	558.2±75.8	559.3±58.8	554.9±83.6	569.0±59.9

Replication Cohort				
Phenotype	ALL (n= 842)	RRMS (n=159)	SPMS (n=536)	PPMS (n=147)
Gender (M/F)	329/513	45/114	220/316	64/83
Age (years)	46.2± 9.6	39.1±10.1	48.4±8.4	46.1±9.4
EDSS (median, range)	5.5 [0-8]	2.5 [0-8]	6[3-7.5]	4.5[2-6.5]
SDMT score (mean, SD)	43.8±14.3	48.4±15.9	43.3±14.0	NA
Disease duration (years)	9.7±8.1	6.1±8.8	12.6±7.2	3.0±3.7
Lesion load (ml)	27.9±27.7	19.5±19.5	30.9±26.2	26.7±37.0
Whole-brain GM volume (ml)	560.2±71.7	565.1±54.4	560.9±78.7	552.6±60.6

Acronyms: EDSS, Expanded Disability Status Scale; M, male, F, female, SDMT, Symbol Digit Modalities Test; NA, Not Available

Comparison of WM and GM sT1w/T2w maps across clinical MS phenotypes

The loading factor of most of the identified WM and GM networks differed at baseline across clinical MS phenotypes. The loading of GM patterns mostly differed between participants with SPMS and RRMS, and between participants with SPMS and PPMS (e.g. GM-IC1). The loading factor of two WM networks (WM-IC1 and WM-IC2) differed between people with RRMS and PPMS, and between people with SPMS and PPMS. The loading factor of three two networks (WM-IC2, WM-IC4, and WM-IC8) differed between people with RRMS and SPMS, and between people with SPMS and PPMS (Supplementary Figure s1).

Stability of WM and GM sT1w/T2w networks across cohorts

Spatial cross-correlations showed that 10 WM and 10 GM networks were stable across cohorts (respectively, $0.40<r<0.75$ and $0.20<r<0.69$) (Supplementary Figure s2). After a visual inspection, 8 WM and 7 GM networks resembled well-known functional and anatomical systems (Figure 2). As shown in Figure 2, ICs encompassed known anatomical regions affected by MS. For instance, WM-IC2 represents a WM sensorimotor network; WM-IC7 the corpus callosum; GM-IC2 was a temporal lobe component, and GM-IC5 an extended DGM component (spanned the thalamus, putamen, caudate, accumbens, frontal and temporal cortex). For a complete description of WM and GM networks, see Figure 2 and Supplementary Materials Table s2.

Relationship between WM and GM sT1w/T2w network measures with GM and WM volume and lesion load

Most of the identified WM and GM sT1w/T2w network measures showed a positive correlation with volumes (e.g., WM-IC4: r= 0.28, 95%CI [0.22:0.34], P<0.0001; GM-IC1: r= 0.15, 95%CI[0.08:0.21], P<0.0001), and a negative correlation with lesion load (e.g., WM-IC4: r= -0.50, 95%CI[-0.55:-0.44], P<0.0001; GM-IC1= r= -0.48, 95%CI[-0.53:-0.43], P<0.0001). Lower WM and GM sT1/T2 network measures (higher microstructural damage) are associated to lower volume measures and with higher lesion load (see Table s3 in Supplementary materials).

Predicting disability progression with Cox regression models

Predicting the risk of 24-week confirmed EDSS progression

Entire discovery and replication cohorts

By the end of the observation period (mean-time-to-progression of 2.07 (SD=1.4) years), 21% (173 out of 842) of participants from the discovery cohort had 24-week CDP. In the replication cohort, after a mean-time-to-progression of 1.9 (SD=1.4) years, 22% (186 out of 842) had 24-week CDP.

The loading factor in the anterior corona radiata (WM-IC1) and temporo-parieto-frontal (GM-IC7) patterns predicted CDP both in the discovery and replication cohorts (WM-IC1 from discovery cohort: hazard ratio [HR]=0.85, 95%CI[0.73:0.99], P<0.05; WM-IC1 from replication cohort: HR=0.84, 95%CI[0.73:0.97], P<0.05; GM-IC7 from discovery cohort: HR=0.83, 95%CI[0.70:0.98], P<0.05; GM-IC7 from replication cohort: HR=0.80, 95%CI[0.69:0.92], P<0.005). After correcting for multiple comparison, GM-IC7 was still statistically significant in the replication cohort (Figure 3 and Supplementary Table s4).

Additionally, in the replication cohort CDP was also predicted by the loading factor in the arcuate fasciculus (WM-IC3), corpus callosum regions (WM-IC7), and DGM (GM-IC5) components.

Lesion load and the volume of whole-brain GM, whole-DGM, and regional volumes of the thalamus, pallidum, and caudate did not predict CDP in the whole discovery and replication cohorts (Figure 3 and Supplementary Table s4).

A post-hoc stepwise backward regression analysis showed that among lesion load, whole brain GM volume, and the variables that were consistently associated to CDP in the two cohorts (WM-IC1 and GM-IC7), the best model to explain CDP included WM-IC1, GM-IC7, and whole brain GM (C-index= 0.62 (se= 0.02)).

Predicting CDP in the RRMS subgroup from the discovery and replication cohorts

When looking within MS phenotypes, we found that different brain networks could predict the 24-week CDP. From the discovery cohort, 22% of patients with RRMS (35 out of 159) experienced CDP during follow-up with a mean-time-to-progression of 2.73 (SD=2.34) years. From the replication cohort, 19.5% (31 out of 159) of participants with RRMS had 24-week CDP by the end of the study, after a mean-time-to-progression of 2.76 (SD=2.40) years.

The loading factor in a sensorimotor pattern (WM-IC2) was consistently associated with a higher risk of developing CDP both in the discovery and replication cohorts (respectively, HR= 0.26, CI[0.14:0.48], P<0.0001; HR=0.51, 95%CI[0.27:0.97], P<0.05) (Figure 3 and

Supplementary Table s4). However, when correcting for multiple components, WM-IC2 was not statistically significant in the replication cohort.

Additionally, the loading factor of a WM arcuate fasciculus [WM-IC3] component, and three GM components (anterior cingulate-precuneus-cerebellum [GM-IC1], cerebellum [GM-IC4], and DGM [GM-IC5] networks) predicted 24-week CDP in the discovery cohort. When correcting for multiple comparisons, GM-IC5 was still statistically significant (Figure 3 and Supplementary Table s4).
In the replication cohort, CDP was also predicted by the loading factor in the anterior corona radiata (WM-IC1) and a parieto-cerebellar (WM-IC2) components.

Lesion load was associated with CDP in the RRMS group from the discovery cohort (HR=1.40, 95%CI [1.00:1.95], P<0.05). However, when correcting for multiple comparisons it was not statistically significant (Figure 3 and Supplementary Table s4). The global volume of whole-brain GM and whole-DGM, and regional volume measures of the thalamus, pallidum, and caudate did not predict CDP in the RRMS cohorts.

Predicting CDP in the SPMS subgroup from the discovery and replication cohorts

In SPMS group in the discovery cohort, 18% (99 out of 537) of participants had 24-week CDP by the end of the study (mean-time-to-progression of 1.98 (SD=1.04). From the replication cohort, 21% (114 out of 536) of participants had CDP within the period of observation, after a mean time-to-progression of 1.81 (SD=1.02) years.

In the discovery cohort, each unit decrease in the loading factor of GM fronto-occipital-somatosensory and motor-cerebellar (GM-IC3) component was associated with a 27% higher risk of developing CDP (HR=0.73, 95%CI [0.57:0.95], P<0.05).
In the replication cohort, for each unit decrease in the loading factor of an anterior corona radiata pattern (WM-IC1) and of a temporo-parieto-frontal component (GM-IC7) there was respectively a 17% and 24% increased risk of developing CDP (HR=0.83, 95%CI[0.70:0.99], P<0.05; HR=0.76, 95%CI[0.64:0.91], P<0.005). When correcting for multiple comparison, GM-IC7 was still statistically significant in the replication cohort (P<0.05).

Lesion load and the global (i.e. whole-brain GM and whole-DGM) and regional (i.e. thalamus, pallidum, and caudate) volume measures did not predict CDP in the SPMS group from the discovery and replication cohorts (Figure 3 and Supplementary Table s4).

Predicting CDP in the PPMS subgroup from the discovery and replication cohorts

In PPMS, 27% (39 out of 147) of participants from the discovery cohort experienced 24-week CDP by end of the study, with a mean-time-to-CDP of 1.67 (SD=1.02) years. For the replication cohort, by end of the observation period, 28% of participants (41 out of 147) had a 24-week CDP (mean-time-to-progression of 1.73 (SD=0.82) years).

In the discovery cohort, the loading factor of a temporo-parieto-frontal component was predictive of CDP (HR=0.51; 95%CI[0.32:0.81]; P<0.005), and was still statistically significant when correcting for multiple comparison. In the replication cohort, the loading factor of a sensorimotor component (WM-IC2) was predictive of CDP (HR=0.44, 95%CI[0.24:0.82], P<0.01), and was still statistically significant when correcting for multiple comparison.

Lesion load and the volume of whole-brain GM, DGM, thalamus, pallidum, and caudate did not predict CDP in the PPMS group from the discovery and replication cohorts (Figure 3 and Supplementary Table s4).

Predicting SDMT worsening

Entire discovery and replication cohorts

In the discovery cohort, longitudinal SDMT data were available for 625 (207 males, 418 females) participants. After a mean time to SDMT worsening of 1.42 (SD=0.08) years, 44% of participants (276 out of 625) experienced a worsening in SDMT by the last available visit. For the replication cohort, SDMT data were available for 624 participants (248 males, 376 females). Of them, 44% of participants (273 out of 624) experienced SDMT worsening by the end of the study, with a mean-time-to-progression of 1.4 (SD=1.06) years.

The loading factor of a WM arcuate fasciculus (WM-IC3), corpus callosum regions (WM-IC7), DGM (GM-IC5), and parieto-cingulate-precuneus-cerebellar (GM-IC6) patterns, and lesion load were consistently associated with a higher risk of developing SDMT worsening both in the discovery and replication cohorts. When correcting for multiple comparison, they were still statistically significant in the replication cohort (Table s5 and Figure 4).

Additionally, in the discovery cohort the loading factor of WM-IC1 (HR=0.85; 95%CI[0.75:0.96]; p<0.01), WM-IC2 (HR=0.74; 95%CI[0.64:0.86]; p<0.0001), and GM-IC3 (HR=0.72; 95%CI[0.62:0.84]; p<0.0001) were associated with SDMT worsening. In the replication cohort, one GM component (GM-IC2= HR=0.87; 95%CI[0.78:0.98]; p<0.05) was associated with SDMT worsening.

Lesion load was consistently associated to SDMT worsening both in the discovery and replication cohorts (respectively, HR=1.27; 95%CI[1.11:1.44]; p<0.0001; and HR=1.26; 95%CI[1.11:1.42]; p<0.0001). The volume of the thalamus predicted SDMT in the replication cohort (HR=0.77; 95%CI[0.66:0.9]; p<0.001), while global measure of whole brain GM or DGM, and regional measures (i.e., volume of caudate and pallidum) were not associated with SDMT worsening (Figure 4 and Supplementary Table s5).
When correcting for multiple comparisons WM-IC3, GM-IC5, and lesion load, and the volume of the thalamus were still statistically significant in both the discovery and replication cohorts.

A post-hoc multivariate stepwise regression analysis showed that among conventional MRI measures and those consistently identified both in the discovery and replication cohort using univariate models, the best model to explain SDMT worsening included GM-IC5 and GM-IC6 (C-index= 0.58, se= 0.017).

Predicting SDMT worsening in the RRMS subgroup from the discovery and replication cohorts

In RRMS, 41.5% of participants (37 out of 89) from the discovery cohort had SDMT worsening by the end of the study (mean-time-to-worsening of 1.14 [SD=0.72] years). For the replication cohort, by the end of the study, 33% (30 out of 90) experienced SDMT worsening (mean-time-to-worsening of 1.22(SD=0.70) years).

In the discovery cohort, the loading factor of a GM temporal component was associated with a higher risk of cognitive worsening (GM-IC2, HR=0.61; 95%CI[0.37:0.99]; p<0.05).
For the replication cohort, the SDMT worsening was predicted by the loading factor in a WM cerebellum network (WM-IC5), and a GM anterior cingulate-precuneus-cerebellum component (GM-IC1) (WM-IC5: HR=0.45; 95%CI[0.23:0.88]; P<0.05; GM-IC1: HR=0.36; 95%CI[0.15:0.87]; p<0.05). However, when correcting for multiple comparison nothing stayed significant (Figure 4 and Supplementary Table s5).

Lesion load, the volume of whole brain GM, of the DGM, and regional volumes measures were not associated with the SDMT worsening.

Predicting SDMT worsening in the SPMS subgroup from the discovery and replication cohorts

In SPMS, 44.5% of participants (238 out of 535) from the discovery cohort had cognitive worsening by the end of the study, with a mean-time-to-worsening of 1.47 (SD=1.13). In the replication cohort, 45.5% of participants (243 out of 534) had SDMT worsening by the end of the study, with a mean-time-to-worsening of 1.37 (SD=1.11) years.

The loading factor of a WM arcuate fasciculus (WM-IC3), caudate nucleus and internal capsule (WM-IC4), and cerebellum (WM-IC5) networks were consistently associated to SDMT worsening both in the discovery and replication cohorts. Each unit increase in these networks was associated to a 18%, 26%, and 21% increased risk of SDMT worsening. When correcting for multiple comparison, the loading factor of WM-IC4 and WM-IC5 were still significant in the discovery cohort (Figure 4 and Table s5).

Additionally, the loading factor of an anterior corona radiata (WM-IC1), anterior cingulate-precuneus-cerebellum (GM-IC1), and fronto-occipital-somatosensory and motor-cerebellar (GM-IC3) components were associated to SDMT worsening (WM-IC1: HR=0.85; 95%CI[0.74:0.98]; p<0.05; GM-IC1: HR=0.84; 95%CI[0.72:0.99]; p<0.05; GM-IC3: HR=0.79; 95%CI[0.65:0.97]; p<0.05). In the replication cohort, the loading factor of a corpus callosum regions (WM-IC7), GM temporal (GM-IC2), and a DGM (GM-IC5) components

were associated with cognitive worsening (WM-IC7: HR=0.75; 95%CI[0.65:0.88]; p<0.0001; GM-IC2: HR=0.88; 95%CI[0.77:1]; p<0.05; and GM-IC5: HR=0.81; 95%CI[0.67:0.99]; p<0.05) (Figure 4 and Table s5).

Lesion load and the volume of the thalamus were consistently associated with SDMT worsening both in the discovery and replication cohorts, while the volume of the whole GM, DGM, and the regional volume measure of caudate and pallidum did not predict SDMT worsening. When correcting for multiple comparison, the loading factor of WM-IC5 and lesion load were still significant in both in the discovery and replication cohorts. (Figure 4 and Table s5).

Discussion

In this study, we found that networks of microstructural changes predict clinical disability and cognitive worsening. Our study is the first to use networks of changes in standardised T1w/T2w in WM and GM to do this. We found that different WM and GM networks predicted clinical progression in the different MS phenotypes. Because sT1w/T2w maps were estimated from MRI scans routinely acquired in clinical trials, our models have potential to be widely applied to future clinical trials, identifying participants at higher risk of progression who are more likely to reveal therapeutic effects in treatment trials.

In our previous work we identified structural networks of covarying GM volumes that predicted SDMT and 9HPT worsening in SPMS using volumetric measures alone, but not confirmed EDSS progression or sT1/T2 ratio.[27] This study adds to our previous work by investigating (1) both WM and GM brain networks, (2) across different MS phenotypes, and (3) by revealing brain networks that are associated with confirmed EDSS progression.

ICA and brain networks

ICA is a data-driven technique used to identify (structural and functional) networks. Spatial-ICA detects modes of covariation in the data; thus, in structural MRI, it identifies networks of

voxels whose values change similarly, although not necessarily in the same direction. Here, we detected WM and GM patterns showing a positive correlation with brain volume measures, suggesting that the higher the T1/T2 ratio (lower microstructural damage), the higher the volume content. Similarly, network measures were inversely correlated with lesion load, suggesting that the higher the T1/T2 ratio, the lower the lesion load.

The identified WM and GM networks resemble well-known anatomical regions, WM tracts, and brain networks. For instance, WM-IC2 represents a WM sensorimotor network; and GM-IC5 is an extended DGM component involving primarily the thalamus, putamen, caudate, cerebellum, frontal and temporal cortex. While several ICA networks appear to represent recognizable anatomical regions rather than distributed neural networks, several components such as GM-IC6 and GM-IC7 represent a cluster of regions (although anatomically close). ICA helps to extract such local by reducing the noise (extracting signals of interest) but also extracting structural networks that would not otherwise be clustered and analysed together. In addition to identifying changes across functional or structural networks, covarying regional changes may also represent disease effects targeting different discrete regions simultaneously or some regions being more (or less) vulnerable to widespread disease effects.[45] Other factors that may contribute to demyelination and neuroaxonal loss leading to brain damage (e.g. local microglial activation, glutamate excitotoxicity, and oxidative injury) might also explain the less conventionally identified networks of covarying microstructural damage.[46]

We found that a cerebellum GM pattern, and six WM networks were similarly identified in the MS and HC cohorts. Results suggest that while some networks might represent more physiological processes, other might be more disease-specific.

WM and GM networks of sT1/T2w changes predict clinical disability and cognitive dysfunction

Among the identified networks, the anterior corona radiata (WM-IC1) and GM temporo-parieto-frontal (GM-IC7) components could predict CDP in the discovery and replication cohorts. The anterior corona radiata and temporo-parieto-frontal networks have extensive afferent and efferent connection, and thus subserve different cognitive and motor functions. The anterior corona radiata is a WM structure associated with the corticopontine, corticobulbar, and corticospinal tracts, with afferent and efferent projections fibers to the cortex. Being this

structure highly involved in motor functions. This might explain the association of this network with EDSS in MS. Previous studies showed an association between higher EDSS and a decrease in T1/T2 ratio values in GM.[20] The predictive ability of the GM network for CDP might be explained by the presence in this component of the parietal cortex. Parietal brain regions have extensive connections with visual, auditory, and somatosensory systems, and sends outputs to several cortical and subcortical areas, particularly to the frontal motor cortex. It plays a crucial role in controlling attention, learning motor skills, planned movements, and proprioception. Damage to this area has already been associated with ataxia and hemispatial neglect.[47] The present results suggest that disability emerges because of a combination of impaired connections between, and damage within brain regions. [20]

The prognostic markers with higher hazard ratios for a clinically meaningful cognitive worsening were a DGM (GM-IC5), a fronto-parieto-cingulate-precuneus-cerebellar (GM-IC6) and a corpus callosum (WM-IC7) patterns. This finding extends our previous work, in which we predicted cognitive worsening using GM volume patterns and identified a GM component spanning similar areas to GM-IC5 with similar predictive performance.[27] The presence of the interconnected basal ganglia nuclei in this component might explain its predictive value. Basal ganglia play a crucial role in controlling motor and non-motor behaviours (such as procedural learning and higher-order process of movement initiation) required by the SDMT task. SDMT is also thought to reflect cognitive fatigue. Previous studies showed that a disconnection syndrome and damage to frontal cortico-subcortical connections could explain the fatigue.[38,48]

The value of sT1/T2w ICA measures in addition to conventional MRI measures

While whole-brain and regional GM volume measures were not associated with clinical disability in the whole, RRMS, SPMS, and PPMS groups from the discovery or replication cohorts, ICA-derived GM and WM components could predict CDP. CDP was also predicted by lesion load in the RRMS in the discovery cohort, although this was not replicated in the replication cohort, and CDP was not predicted by lesion load in either PPMS or SPMS.

Considering cognitive worsening, while WM-IC5, WM-IC7, GM-IC5, GM-IC6, and lesion load could consistently predict SDMT worsening both in the discovery and replication cohorts, whole-brain GM volume measures could not. When performing post-hoc backward selection-models, we found that among conventional MRI measures and those that were consistently predictive CDP in both cohorts in the univariate models, the final model included an anterior corona radiata, a GM temporo-parieto-frontal network and whole brain GM volume. For cognitive worsening, the final model included a deep GM and a parieto-cingulate-precuneus-cerebellar network. For cognitive worsening, the final model included a deep GM and a parieto-cingulate-precuneus-cerebellar network. A previous study has shown that decreased T1/T2 ratio values in the cingulate cortex were associated with the attentional performance in patients with MS. [16] Therefore, the present results suggest that T1w/T2w network measures may complement conventional MRI measures in predicting clinical progression and cognitive worsening.

Overall a reduction in T1/T2 seems to be a proxy for pathological changes in MS.[10,11,13] However, it is important to note that T1/T2 is a not a purely quantitative measures, and in MS similar to other neurodegenerative disorders, it can change as a result of many different processes that may change as T1/T2 changes in different directions. These changes may be due to myelin, neurite density, and iron deposition. A recent study showed that in the cortex of Alzheimer's disease patients the T1/T2 index increased compared to healthy controls. While the pathological processes in MS are different, these results show that T1/T2 changes should be interpreted with caution.[49]

Strengths and limitations

Computational limits meant that we could only process data from 1900 participants. By using data from eight different clinical trials, and dividing the cohort in two, we were able to determine if the results were robust to confounding by MRI protocols and scanners, and if they could be replicated across cohorts. An alternative approach would have been to use data from fewer trials and fewer scanners, so reducing heterogeneity in the MRI data, and this may increase sensitivity to more subtle patterns of changes, but conversely could reduce confidence that the results were robust to protocol and scanner heterogeneity that is likely to be encountered in future multicentre clinical trials. We included in this study both 2D and 3D

images, acquired at 1.5 or 3T. While T1/T2 ratio values were not equal when comparing between MRI protocols, these groups also differed for age, sex, disease duration, EDSS score and clinical phenotypes. This should not lead to spurious ICA patterns being identified, but it could obscure them, and while MRI protocol was used as a covariate in the statistical analyses, associations with clinical features may have been influenced by this.When compared with using volumetric 3D T1-weighted scans alone, the use of 3D and 2D images may have reduced sensitivity to more subtle covarying patterns, and so we may have overlooked networks that are clinically relevant but more difficult to detect.

Similarly, in this study, we did not use DTI or MTR data that are more sensitive to intrinsic tissue changes, and further work using such data may reveal further clinically relevant networks affected by MS. However, these measures are not routinely collected in phase 3 clinical trials, while T1w and T2w MRI sequences normally are, and so the present findings can be more readily translated.

In this study we used NAGM as a scaling factor to obtain sT1/T2 measures which is not ideal. There is no gold-standard on the calibration technique that should be preferred to obtain the standardized T1/T2 ratio maps. Several techniques are available, although they all have pitfalls. We found that the technique used in this study was consistent with other calibration techniques[9,39] (in GM: ICC 0.70; in WM: ICC 0.79). We preferred this method over using extra-CNS structures (e.g. eyes and temporal muscles) as it would not be applicable to anonymized and defaced MRI data where information for eyes and temporal muscles is missing. Moreover, the temporal muscle is a thin structure and there are no automatic techniques to segment it. Therefore, it needs to be manually drawn from the MNI template and registered to the subject space, requiring a certain level of manual editing when the registration was suboptimal which is not feasible for a large cohort. Finally, using GM should not introduce spurious regional differences, but it may reduce the magnitude of effects (i.e. potential bias will more likely be in the conservative direction).

Our cohort mainly included participants with progressive forms of MS. Future studies should investigate whether when repeating the analysis on a large cohort of participants with RRMS other components might emerge. Additionally, Comparing HC (from the HCP) with the MS groups, we found that those with MS had lower GM and WM sT1w/T2w overall, and some components seen in the MS groups were also seen in HC, suggesting that they may be part of

a structural response to brain pathology per se, rather than MS specific. For this study we used data from the HCP as no HC data were available in phase 3 clinical trials, and given that the HCP data cannot be matched to the clinical trial data in terms of the scanners and protocols used, comparisons should be considered with caution and future studies in age and MRI protocol matched cohort of HC and MS participants are needed to confirm how MS-specific each component is."

We showed that 2 networks measures and 4 components were consistently associated with CDP and cognitive worsening in the whole discovery and replication cohorts, results were not completely replicated when splitting the cohorts into clinical phenotypes. The discrepancy across two data sets is because they represent different population of patients at different stages of the disease, different eligibility criteria in the clinical trials, and scanner effects. The primary aim of our study was to find the networks that remain robust across cohorts and, despite all the differences mentioned before, our identified measure can reliably be used in future clinical trials.

Our cohorts had a relatively short follow-up (average years 2.75 [1.74]), consistent with the typical length of clinical trials and our interest in applications in future clinical trials. Future work will apply these methods in observational cohorts which allow for longer follow ups to investigate longer term outcomes in MS patients.

EDSS was the only clinical data that was available across all clinical trials, while 9 Hole Peg Test (9HPT) and Timed 25-Foot Walk Test (T25FW) data might have allowed to investigate more specifically the upper and lower body functions notoriously impaired in MS. Similarly, cognitive data were limited to SDMT, and more comprehensive batteries might allow investigating the association of these patterns with other cognitive domains known to be impaired in MS. We used SDMT because it is believed to be relevant to information processing speed that is the most impaired cognitive function in MS and it has recently been validated as a sensitive measure of cognitive impairment.[38]

Conclusion

We found clinically relevant GM and WM networks of microstructural changes from sT1w/T2w WM and GM maps, some of which predicted the disability progression and

cognitive worsening in a large cohort of participants with MS. Further work is needed to resolve the underlying pathological processes leading to these network changes, and their independence from whole-brain atrophy and lesions. The identification of networks of GM and WM changes, based on MRI scans routinely collected in clinical trials, that predict progression may be used to identify treatment trial participants most likely to experience disability progression.

Acknowledgments

This investigation was supported (in part) by (an) award(s) from the International Progressive MS Alliance, award reference number PA-1412-02420. We are grateful to all the IPMSA investigators who have contributed trial data to this study as part of EPITOME: Enhancing Power of Intervention Trials Through Optimized MRI Endpoints network. DC, FB, FP and OC are supported by the NIHR biomedical research centre at UCLH.

We are grateful to all the MS-SMART investigators. MS-SMART was an investigator-led project sponsored by University College London (UCL), Edinburgh Clinical Trials Unit (ECTU), QSMSC NMR/MRI analysis centre, NIHR Efficacy and Mechanism Evaluation (EME) Programme, MS Clinical Trials Network (MS CTN), the UK Multiple Sclerosis (MS) Society and the US National MS Society.

Data collection and sharing for part of this project was provided by the Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D.,Van J. Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of Southern California.

Author Contributions

E.C., D.T.C., and A.E. contributed to study concept and design.
N.S., D.L.A., F.B., O.C., D.T.C., and A.E. contributed to data acquisition.
E.C., F.P., J.S., and A.E. contributed to data analysis.
E.C., A.B., F.B., O.C., D.T.C., and A.E. contributed to results interpretation.
E.C., F.P., J.S., N.S., D.L.A., C.G.A.M.W., F.B., O.C., D.T.C., and A.E. contributed to drafting of the manuscript and figures.

Potential Conflicts of Interest

The authors have no competing interests with respect to this research. The full disclosure statement is as follows:

DLA reports consulting fees from Albert Charitable Trust, Alexion Pharma, Biogen, Celgene, Frequency Therapautics, Genentech, Med-Ex Learning, Merck, Novartis, Population Council, Receptos, Roche, and Sanofi-Aventis, grants from Biogen, Immunotec and Novartis, and an equity interest in NeuroRx. F.B has received compensation for consulting services and/or speaking activities from Bayer Schering Pharma, Biogen Idec, Merck Serono, Novartis, Genzyme, Synthon BV, Roche, Teva, Jansen research and IXICO and is supported by the NIHR Biomedical Research Centre at UCLH. OC has received research grants from the MS Society of Great Britain & Northern Ireland, National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre, EUH2020, Spinal Cord Research Foundation, and Rosetrees Trust. She serves as a consultant for Novartis, Teva, and Roche and has received an honorarium from the American Academy of Neurology as Associate Editor of Neurology and serves on the Editorial Board of Multiple Sclerosis Journal. DC is a consultant for Hoffmann-La Roche. In the last three years he has been a consultant for Biogen, has received research funding from Hoffmann-La Roche, the International Progressive MS Alliance, the MS Society, the Medical Research Council, and the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre, and a speaker's honorarium from Novartis. He co-supervises a clinical fellowship at the National Hospital for Neurology and Neurosurgery, London, which is supported by Merck. AE has received speaker's honoraria from Biogen and At The Limits educational programme. AE has received research grants from Medical Research Council, UK Research and Innovation's Innovate UK, Biogen, UCL Innovation and Enterprise, Roche and Merck. AE serves on the Editorial Board of Neurology. He has received travel support from the National Multiple Sclerosis Society and honorarium from the Journal of Neurology, Neurosurgery and Psychiatry for Editorial Commentaries. AE and FB have equity stake in Queen Square Analytics. FP was funded by a Guarantors of Brain non-clinical Postdoctoral Fellowship. EC and JS have nothing to disclose.

References

1. Langdon DW. Cognition in multiple sclerosis. Curr Opin Neurol 2011;24(3):244–249.
2. Gaetano L, Magnusson B, Kindalova P, et al. White matter lesion location correlates with disability in relapsing multiple sclerosis. Mult Scler J Exp Transl Clin 2020;6(1):205521732090684.
3. Chung KK, Altmann D, Barkhof F, et al. A 30-Year Clinical and Magnetic Resonance Imaging Observational Study of Multiple Sclerosis and Clinically Isolated Syndromes. Ann Neurol 2020;87(1):63–74.
4. Fisher E, Lee JC, Nakamura K, Rudick RA. Gray matter atrophy in multiple sclerosis: A longitudinal study. Ann Neurol 2008;64(3):255–265.
5. De Stefano N, Giorgio A, Battaglini M, et al. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. Neurology 2010;74(23):1868–1876.
6. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. Curr Opin Neurol 2002;15(3):239–45.
7. Cooper G, Finke C, Chien C, et al. Standardization of T1W/T2W ratio improves detection of tissue damage in multiple sclerosis. Front Neurol 2019;10(APR):1–10.
8. Rocca MA, Riccitelli GC, Meani A, et al. Cognitive reserve, cognition, and regional brain damage in MS: A 2 -year longitudinal study. Multiple Sclerosis Journal 2019;25(3):372–381.
9. Margoni M, Pagani E, Meani A, et al. Exploring in vivo multiple sclerosis brain microstructural damage through T1w/T2w ratio: a multicentre study [Internet]. J Neurol Neurosurg Psychiatry 2022;93(7)[cited 2022 Jul 24] Available from: https://pubmed.ncbi.nlm.nih.gov/35580993/
10. Hannoun S, Kocevar G, Codjia P, et al. T1/T2 ratio: A quantitative sensitive marker of brain tissue integrity in multiple sclerosis. Journal of Neuroimaging 2022;32(2):328–336.
11. Zheng Y, Dudman J, Chen JT, et al. Sensitivity of T1/T2-weighted ratio in detection of cortical demyelination is similar to magnetization transfer ratio using post-mortem MRI. Multiple Sclerosis Journal 2022;28(2):198–205.
12. Pareto D, Garcia-Vidal A, Alberich M, et al. Ratio of T1-weighted to T2-weighted signal intensity as a measure of tissue integrity: Comparison with magnetization transfer ratio in patients with multiple sclerosis. American Journal of Neuroradiology 2020;41(3):461–463.
13. Nakamura K, Zheng Y, Ontaneda D. T1/T2-weighted ratio is a surrogate marker of demyelination in multiple sclerosis—yes. Multiple Sclerosis Journal 2022;28(3):352–354.
14. Rovira À, Pareto D. T1/T2-weighted ratio is a surrogate marker of demyelination in multiple sclerosis – Commentary. Multiple Sclerosis Journal 2022;28(3):357–358.
15. Mühlau M. T1/T2-weighted ratio is a surrogate marker of demyelination in multiple sclerosis: No. Multiple Sclerosis Journal 2022;28(3):355–356.
16. Righart R, Biberacher V, Jonkman LE, et al. Cortical pathology in multiple sclerosis detected by the T1/T2-weighted ratio from routine magnetic resonance imaging. Ann Neurol 2017;82(4):519–529.
17. Glasser MF, van Essen DC. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. Journal of Neuroscience 2011;31(32):11597–11616.

18. Nakamura K, Chen JT, Ontaneda D, et al. T1-/T2-weighted ratio differs in demyelinated cortex in multiple sclerosis. Ann Neurol 2017;82(4):635–639.
19. Righart R, Biberacher V, Jonkman LE, et al. Cortical pathology in multiple sclerosis detected by the T1/T2-weighted ratio from routine magnetic resonance imaging. Ann Neurol 2017;82(4):519–529.
20. Boaventura M, Sastre-Garriga J, Garcia-Vidal A, et al. T1/T2-weighted ratio in multiple sclerosis: A longitudinal study with clinical associations. Neuroimage Clin 2022;34:102967.
21. Steenwijk MD, Geurts JJG, Daams M, et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain 2016;139(1):115–126.
22. Muthuraman M, Fleischer V, Kroth J, et al. Covarying patterns of white matter lesions and cortical atrophy predict progression in early MS. Neurology(R) neuroimmunology & neuroinflammation 2020;7(3):1–12.
23. Zhang J, Giorgio A, Vinciguerra C, et al. Gray matter atrophy cannot be fully explained by white matter damage in patients with MS. Multiple Sclerosis Journal 2020;1–13.
24. Bergsland N, Horakova D, Dwyer MG, et al. Gray matter atrophy patterns in multiple sclerosis: A 10-year source-based morphometry study. Neuroimage Clin 2018;17(November 2017):444–451.
25. Rocca MA, Valsasina P, Meani A, et al. Association of Gray Matter Atrophy Patterns With Clinical Phenotype and Progression in Multiple Sclerosis. Neurology 2021;96(11):e1561–e1573.
26. Huiskamp M, Eijlers AJC, Broeders TAA, et al. Longitudinal Network Changes and Conversion to Cognitive Impairment in Multiple Sclerosis. Neurology 2021;97(8):e794–e802.
27. Colato E, Stutters J, Tur C, et al. Predicting disability progression and cognitive worsening in multiple sclerosis with gray matter network measures Radiological advances i (NAIMS/MAGNIMS). Multiple Sclerosis Journal 2020;26(3 SUPPL):22–23.
28. Meijer KA, Cercignani M, Muhlert N, et al. Patterns of white matter damage are non-random and associated with cognitive function in secondary progressive multiple sclerosis. Neuroimage Clin 2016;12:123–131.
29. Spain R, Powers K, Murchison C, et al. Lipoic acid in secondary progressive MS. Neurol Neuroimmunol Neuroinflamm 2017;4(5)
30. Montalban X, Hauser SL, Kappos L, et al. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis [Internet]. New England Journal of Medicine 2017;376(3):209–220.Available from: http://www.nejm.org/doi/10.1056/NEJMoa1606468
31. Chataway J, de Angelis F, Connick P, et al. Efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis (MS-SMART): a phase 2b, multiarm, double-blind, randomised placebo-controlled trial. Lancet Neurol 2020;19(3):214–225.
32. Hawker K, O'Connor P, Freedman MS, et al. Rituximab in patients with primary progressive multiple sclerosis: Results of a randomized double-blind placebo-controlled multicenter trial. Ann Neurol 2009;66(4):460–471.
33. Gold R, Kappos L, Arnold DL, et al. Placebo-Controlled Phase 3 Study of Oral BG-12 for Relapsing Multiple Sclerosis. New England Journal of Medicine 2012;367(12):1098–1107.
34. Hauser SL, Bar-Or A, Comi G, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. New England Journal of Medicine 2017;376(3):221–234.

35. Kapoor R, Ho PR, Campbell N, et al. Effect of natalizumab on disease progression in secondary progressive multiple sclerosis (ASCEND): a phase 3, randomised, double-blind, placebo-controlled trial with an open-label extension. Lancet Neurol 2018;17(5):405–415.
36. Chataway J, De Angelis F, Connick P, et al. Efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis (MS-SMART): a phase 2b, multiarm, double-blind, randomised placebo-controlled trial. Lancet Neurol 2020;19(3):214–225.
37. Tur C, Moccia M, Barkhof F, et al. Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. Nat Rev Neurol 2018;14(2):75–93.
38. Benedict RHB, Tomic D, Cree BA, et al. Siponimod and Cognition in Secondary Progressive Multiple Sclerosis: EXPAND Secondary Analyses. Neurology 2021;96(3):e376–e386.
39. Ganzetti M, Wenderoth N, Mantini D. Whole brain myelin mapping using T1- and T2-weighted MR imaging data. Front Hum Neurosci 2014;8(SEP)
40. Tustison NJ, Avants BB, Cook PA, et al. N4ITK : Improved N3 Bias Correction. 2010;1310–1320.
41. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61–78.
42. Cardoso MJ, Modat M, Wolz R, et al. Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. IEEE Trans Med Imaging 2015;34(9):1976–1988.
43. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. JMLR 2011;12:2825–2830.
44. Esposito F, Scarabino T, Hyvarinen A, et al. Independent component analysis of fMRI group studies by self-organizing clustering. Neuroimage 2005;25(1):193–205.
45. Chard DT, Miller DH. What lies beneath grey matter atrophy in multiple sclerosis? Brain 2016;139(1):7–10.
46. Popescu V, Klaver R, Voorn P, et al. What drives MRI-measured cortical atrophy in multiple sclerosis? Multiple Sclerosis 2015;21(10):1280–1290.
47. Mattingley JB, Husain M, Rorden C, et al. Motor role of human inferior parietal lobe revealed in unilateral neglect patients. Nature 1998 392:6672 1998;392(6672):179–182.
48. Filippi M, Preziosa P, Rocca MA. Brain mapping in multiple sclerosis: Lessons learned about the human brain. Neuroimage 2019;190(June 2017):32–45.
49. Pelkmans W, Dicks E, Barkhof F, et al. Gray matter T1-w/T2-w ratios are higher in Alzheimer's disease. Hum Brain Mapp 2019;40(13):3900–3909.
50. Margoni M, Pagani E, Meani A, et al. Exploring in vivo multiple sclerosis brain microstructural damage through T1w/T2w ratio: a multicentre study. J Neurol Neurosurg Psychiatry 2022;93(7):741–752.

List of Figures

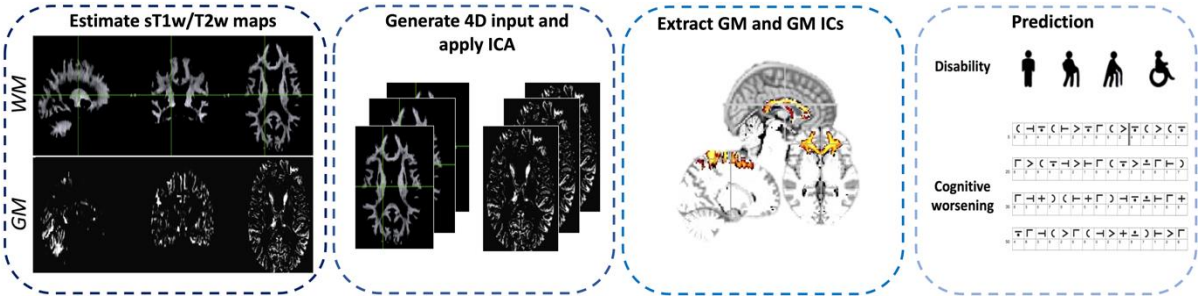
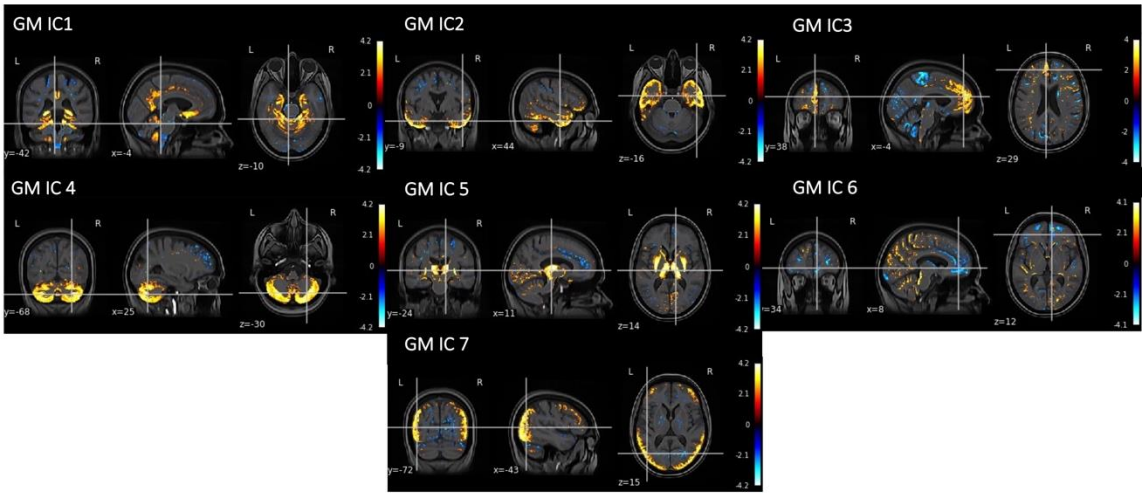
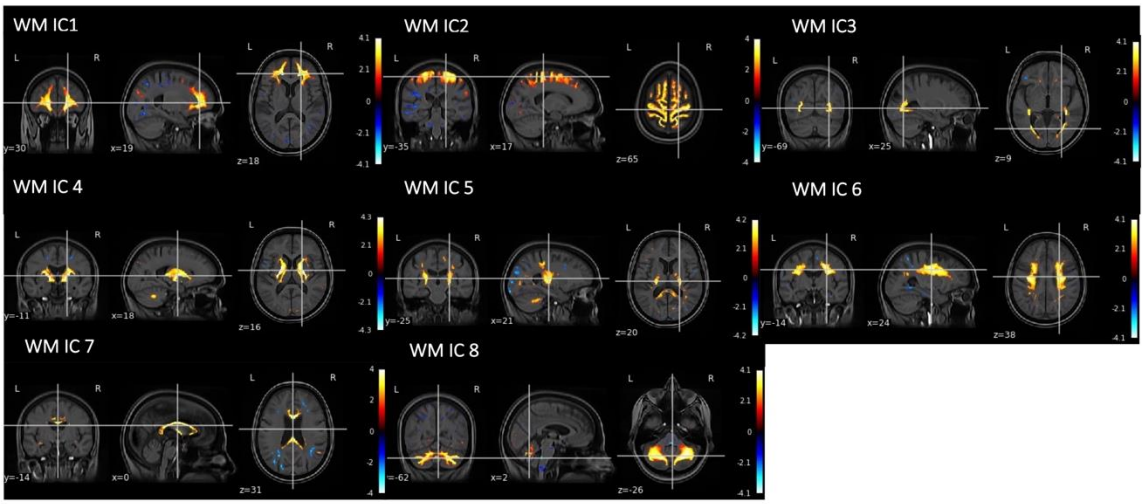


Figure 1 Pipeline to obtain networks of sT1w/T2w changes and determine their predictive value. We first split the data into discovery and replication cohort; estimated WM and GM sT1w/T2w maps; generated a WM and a GM 4D image used as input of the ICA; identified GM and WM patterns of covarying sT1w/T2w changes; determined the stability of the identified components and their prognostic value for disability progression and cognitive worsening.

Acronyms: sT1w/T2w, standardized T1w/T2w; WM, white matter; GM, grey matter; ICA, Independent Component Analysis; EDSS, Expanded Disability Status Scale; SDMT, Symbol Digit Modalities Test



A



B

Figure 2 Brain areas that had similar patterns of change in WM and GM sT1w/T2w. Figure 2 shows stable (**A**) GM and (**B**) WM patterns of covarying sT1w/T2w changes identified by the ICA . The colour bar represents the loading of each network: the higher the loading, the lower the microstructural damage in the network.

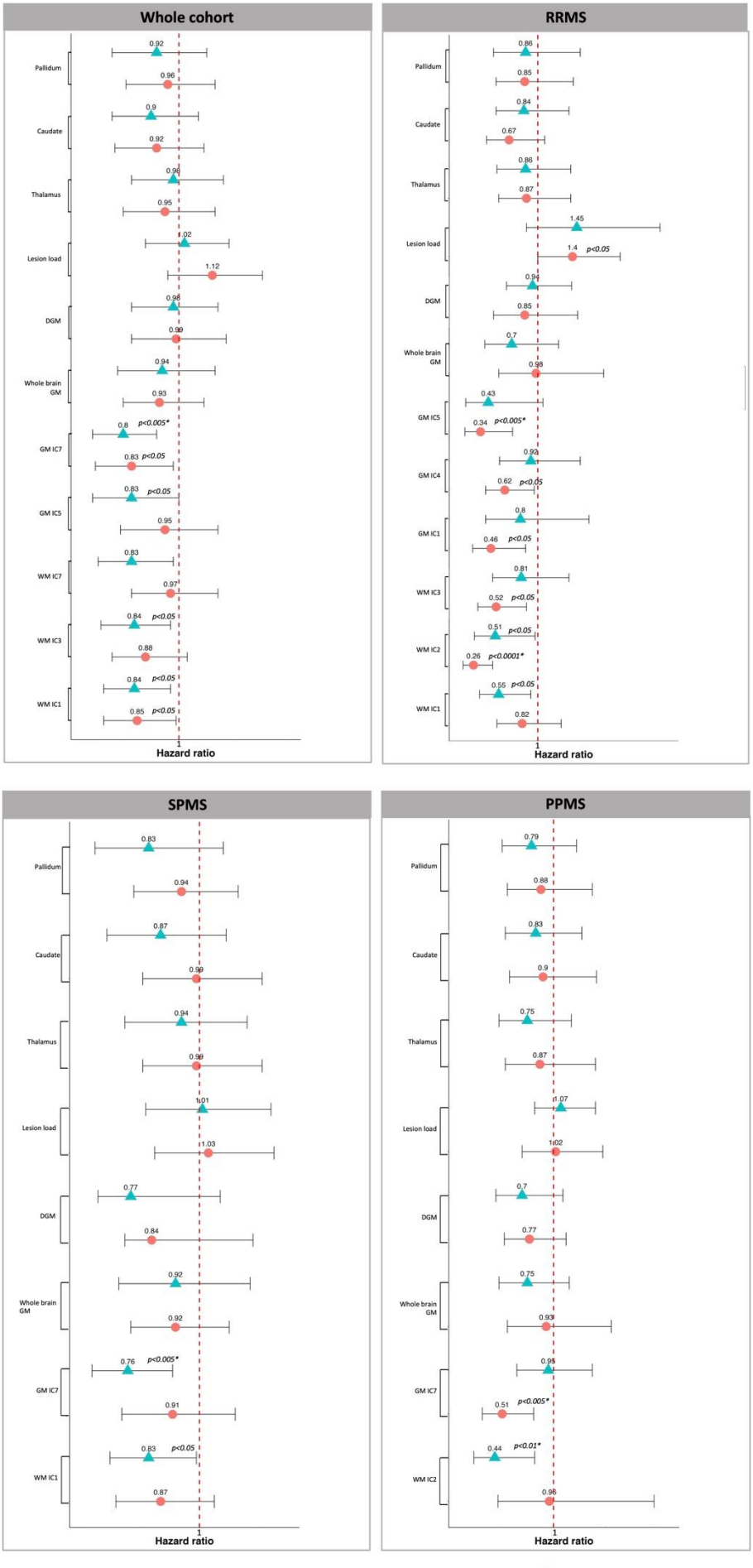

Whole cohort
Pallidum
0.92
0.96
Caudate
0.9
0.92
Thalamus
0.98
0.95
Lesion load
1.02
1.12
DGM
0.98
0.99
Whole brain GM
0.94
0.93
GM IC7
0.8 p<0.005*
0.83 p<0.05
GM IC5
0.83 p<0.05
0.95
WM IC7
0.83
0.97
WM IC3
0.84 p<0.05
0.88
WM IC1
0.84 p<0.05
0.85 p<0.05
1
Hazard ratio
RRMS
Pallidum
0.86
0.85
Caudate
0.84
0.67
Thalamus
0.86
0.87
Lesion load
1.45
1.4 p<0.05
DGM
0.94
0.85
Whole brain GM
0.7
0.98
GM IC5
0.43
0.34 p<0.005*
GM IC4
0.93
0.62 p<0.05
GM IC1
0.8
0.46 p<0.05
WM IC3
0.81
0.52 p<0.05
WM IC2
0.51 p<0.05
0.26 p<0.0001*
WM IC1
0.55 p<0.05
0.82
1
Hazard ratio
SPMS
Pallidum
0.83
0.94
Caudate
0.87
0.99
Thalamus
0.94
0.99
Lesion load
1.01
1.03
DGM
0.77
0.84
Whole brain GM
0.92
0.92
GM IC7
0.76 p<0.005*
0.91
WM IC1
0.83 p<0.05
0.87
1
Hazard ratio
PPMS
Pallidum
0.79
0.88
Caudate
0.83
0.9
Thalamus
0.75
0.87
Lesion load
1.07
1.02
DGM
0.7
0.77
Whole brain GM
0.75
0.93
GM IC7
0.98
0.51 p<0.005*
WM IC2
0.44 p<0.01*
0.96
1
Hazard ratio
Discovery cohort
Replication cohort

Figure 3 Survival analysis for 24-week CDP.

Figure 3 shows the hazard ratio (HR) of the white and grey matter components that were predictive of the 24-week CDP in the discovery or replication cohorts, for the whole cohort, RRMS, SPMS, and PPMS groups. HR higher than 1 indicates that each standard deviation increase in the loading of the corresponding variable is associated to a higher risk of developing the event. HR values lower than 1 indicates that for each standard deviation decrease in the loading of the corresponding variable, there is a higher risk of 24-week CDP.

Error bars represent the confidence interval (CI).

* Statistically significant after Benjamini-Hochberg correction

Acronyms: HR= hazard ratio; EDSS= expanded disability status scale; RRMS= relapsing-remitting MS; SPMS= secondary progressive MS; PPMS= primary progressive MS; WM-IC = white matter independent component; GM-IC = grey matter independent component

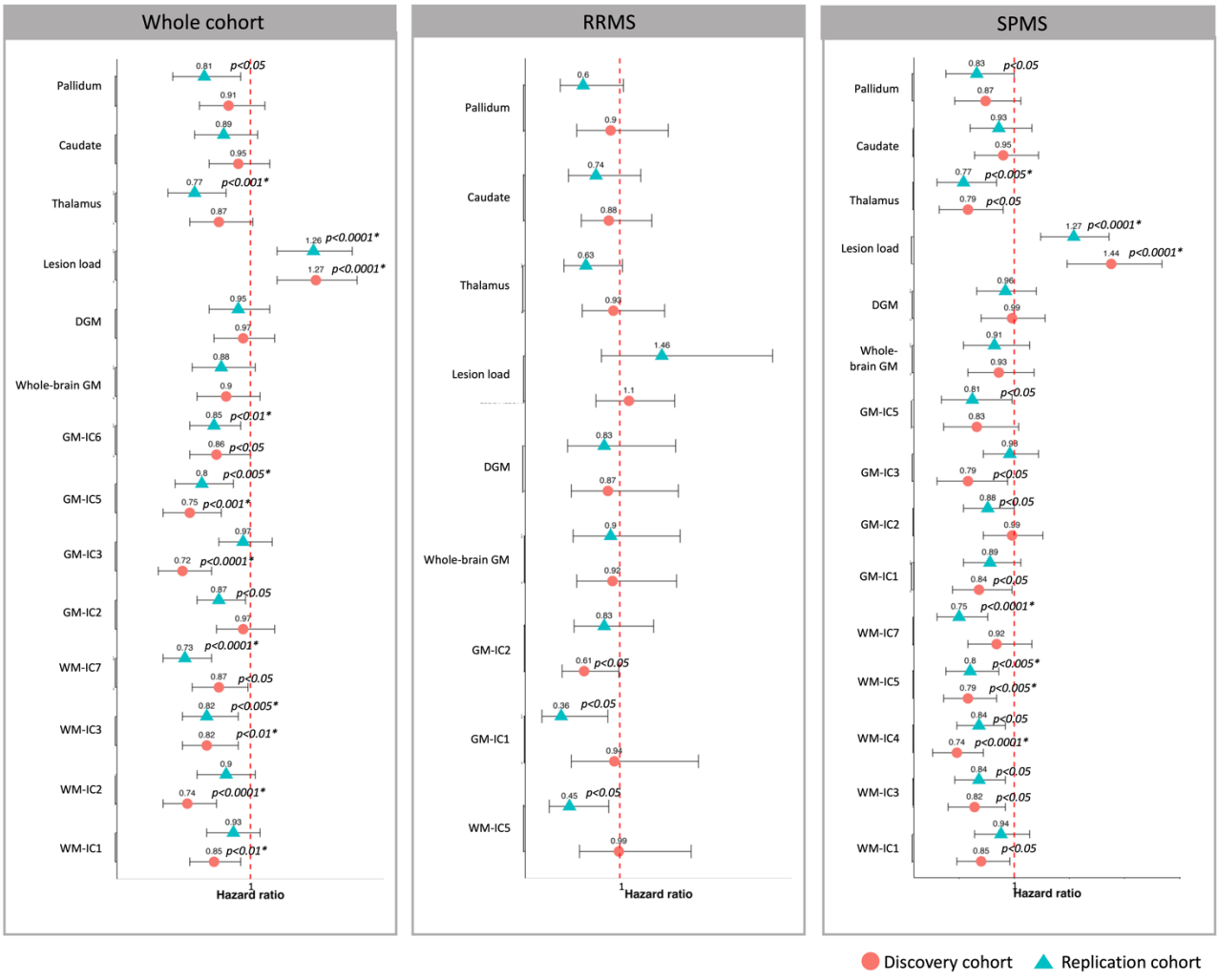


Figure 4 Survival analysis for SDMT worsening.

Figure 4 shows the hazard ratio (HR) of WM and GM components that were predictive of SDMT worsening in the discovery or replication cohorts, for the whole cohort, RRMS, and SPMS groups. HR higher than 1 indicates that for each standard deviation increase in the loading of the corresponding variable there is a higher risk of developing the event. HR values lower than 1 indicates that each standard deviation decrease in the loading of the corresponding variable is associated to a higher risk of developing cognitive worsening. For instance, considering the whole cohorts, for each standard deviation decrease in GM volume there is a 13% and 18% increased risk of developing SDMT worsening in the discovery and replication cohorts. Error bars represent the confidence interval (CI).

* Statistically significant after Benjamini-Hochberg correction

Acronyms: HR= hazard ratio; SDMT= Symbol digit modalities test; WM-IC = white matter independent component; GM-IC = grey matter independent component; RRMS= relapsing-remitting MS; SPMS= secondary progressive MS

1

Supplementary Materials

Acquisition protocol

This was a retrospective study and the MRI protocol for each clinical trial is previously published. We included brain scans acquired at either 1.5 or 3T from 8 clinical trials. For completeness, a brief summary of acquisition protocol for each study is as follows:

- ASCEND:
  - T1-weighted, voxel size= 1x1x3mm^3, ET= 11, RT=30, Flip Angle= 30
  - T2-weighted, voxel size= 1x1x3mm^3, ET= 83, RT= 5470, Flip Angle= 180
- DCE:
  - T1-weighted, voxel size= 1x1x3mm^3, ET= 20, RT= 600, Flip Angle= 90
  - T2-weighted, voxel size= 1x1x3mm^3, ET= 84, RT= 2800, Flip Angle= 90
- LIPOIC-ACID:
  - T1-weighted, voxel size= 1x1x1mm^3, ET= 3.9, RT= 8.45, Flip Angle= 8.0
  - T2-weighted, voxel size= 0.5x0.5x3mm^3, ET= 100, RT= 3585.5, Flip Angle= 90
- MS-SMART:
  - T1-weighted, voxel size= 1x1x1mm^3, ET= 0.004, RT= 2.4, Flip Angle= 1
  - T2-weighted, voxel size= 1x1x3mm^3, ET=0.01, RT= 2.92, Flip Angle= 150
- OLYMPUS:
  - T1-weighted, voxel size= 1x1x3mm^3, ET= 20, RT= 650, Flip Angle= 90
  - T2-weighted, voxel size= 0.4x0.4x3mm^3, ET= 2500, RT= 94.9, Flip Angle= 90
- OPERA1, OPERA2, ORATORIO:
  - T1-weighted, voxel size= 1x1x3mm^3, ET= 0.01, RT= 0.03, Flip Angle= 30
  - T2-weighted, voxel size= 1x1x3mm^3, ET= 0.09, RT= 5.75, Flip Angle= 180

The table below shows the number of participants scanned at 1.5 or 3T, and using 2D or 3D acquisitions.

Clinical trial	N	Field strength (1.5 or 3T)	Acquisition (2D or 3D)
ASCEND	814	1.5	3D
DCE	136	1.5	2D
LIPOIC-ACID	26	3	3D
MS-SMART	269	3	3D
OLYMPUS	85	1.5	2D
OPERA1	95	1.5	3D
OPERA2	61	1.5	3D
ORATORIO	209	1.5	2D

Considering the field strength, 295 participants were scanned at 3T and 1400 participants at 1.5 T. The two groups were not equal for age (mean 54.9 (SD=7) vs. mean 44.7(SD= 9), P<0.0001), disease duration (mean 16 years (SD= 10), vs. mean 8.7 years (SD=7.2), P<0.0001), EDSS score (median 6 [5.5-6.5] vs. median 5.5 [4-6], P<0.0001), sex (P<0.0001), mean GM T1/T2 ratio values (mean 0.07 (SD= 0.01) vs. mean (0.05 (SD = 0.01), P<0.0001), mean WM (mean 0.25 (SD= 0.05) vs. mean 0.17 (SD= 0.03), P<0.0001), and clinical phenotypes (3T group included only participants with SPMS).

Considering the acquisition, 1265 participants were scanned using a 3D acquisition and 430 participants using a 2D acquisition. The two groups were not equal for age (mean 47.3 (SD=9.3) vs. mean 44.0(SD= 9.4), P<0.0001), disease duration (mean 12 years (SD= 8.4), vs. mean 3.9 years (SD=4.3), P<0.0001), EDSS score (median 6 [4.5-6.5] vs. median 4 [3-6], P<0.0001), mean GM T1/T2 ratio values (mean 0.05 (SD= 0.01) vs. mean (0.04 (SD = 0.01), P<0.0001), mean WM (mean 0.19 (SD= 0.05) vs. mean 0.17 (SD= 0.04), P<0.0001), and clinical phenotypes (the 3D acquisition group included mainly participants with SPMS (N=1109) and RRMS (N=156), while the 2D acquisition group included mainly participants with PPMS (N=294) and RRMS (N=136)).

We used the MRI protocol as a covariate in the following statistical analysis.

Standardised T1w/T2w maps

To extract sT1w/T2w maps, we applied a pipeline similar to Cooper *et al.* [1]. Specifically:

1) N4 bias filed corrected T1w and T2w scans to correct for scanner-field inhomogeneity;[2]
2) Rigid registered T1w and T2w images to an halfway space using NiftyReg;[3]
3) Applied an established pipeline as in *Eshaghi et al.*[4] and used Geodesic Information Flows (GIF) version 3.0[5] to obtain segmentation maps and DeepMedic software [6] for lesion masks in the native T1w lesion filled and FLAIR images;
4) Used fslmaths to extract GM, WM and CSF maps from the previous step and NiftyReg to register them to the halfway space. With *fslmaths*, we obtained masks of GM and WM for each participant in the mid-space;
5) Rigid registered FLAIR images to the halfway space and transformed lesion masks from the native FLAIR to the mid-space;
6) To obtain GM maps, we multiplied GM masks by the T1w and T2w N4-bias field corrected images in halfway space and subtracted lesion masks;
7) Estimated the median values in T1w and T2w GM and calculated a scaling factor by dividing median values in GM in T1w by the median values GM in T2w;
8) Obtained a scaled T2w image by multiplying the T2w scan by the estimated scaling factor;
9) To compute standardized T1w/T2w (sT1w/T2w) maps in the halfway space we applied the following formula: sT1w/T2w = T1w – sT2w / T1w + sT2w;
10) To obtain GM and WM sT1w/T2w maps, for each participant we multiplied respectively the whole brain sT1w/T2w maps by their GM and WM masks;
11) We transformed the GM and WM sT1w/T2w maps to a customised template obtained as in *Colato et al.*[7]. To be even more conservative in the definition of the WM maps, we created a WM mask using the major voting algorithm implemented in antsJointLabelFusion to fuse the 39 parcellation maps of subjects that had contributed to the template and obtain a single parcellation map where the likelihood that each region corresponds to the labelled one is higher. [8] We thresholded out non-WM brain structures (e.g., ventricles, meninges, etc.), binarised the output and used the obtained mask to mask WM sT1w/T2w maps.

Standardised T1w/T2w maps and networks in healthy controls

To determine whether similar networks were present in healthy controls (HC), we randomly selected and downloaded the available 3D T1 and T2 scans from 300 participants from the Human Connectome Project.

MRI analysis

ST1w/T2w maps in healthy controls

We adapted the previously described pipeline to healthy controls. Specifically:

1) N4 bias filed corrected T1w and T2w scans to correct for scanner-field inhomogeneity;[2]
2) Used Geodesic Information Flows (GIF) version 3.0[5] to obtain segmentation maps in the native T1w filled images;
3) Used fslmaths to extract GM, WM and CSF maps from the previous step and NiftyReg to register them to the halfway space. With *fslmaths*, we obtained masks of GM and WM for each participant in the mid-space;
4) To obtain GM maps, we multiplied GM masks by the T1w and T2w N4-bias field corrected images in halfway space;
5) Estimated the median values in T1w and T2w GM and calculated a scaling factor by dividing median values in GM in T1w by the median values GM in T2w;
6) Obtained a scaled T2w image by multiplying the T2w scan by the estimated scaling factor;
7) To compute standardized T1w/T2w (sT1w/T2w) maps in the halfway space we applied the following formula: sT1w/T2w = T1w – sT2w / T1w + sT2w;
8) To obtain GM and WM sT1w/T2w maps, for each participant we multiplied respectively the whole brain sT1w/T2w maps by their GM and WM masks;
9) We transformed the GM and WM sT1w/T2w maps to the MNI template and applied a WM and GM masks.

We visually inspected the output of each step to check for segmentation errors (e.g. WM/GM maps estimation), and misregistration.

ST1w/T2w maps in healthy controls

We repeated ICA for HC using the FastICA algorithm from scikit-learn 0.23.1 [34] to identify spatial patterns of covarying 1) WM changes and 2) GM changes from sT1w/T2w ratio maps in the study-specific template. We specified the number of components to be identified to 20 for ICA to capture potentially relevant but less strong data patterns from WM sT1w/T2w maps.[35] We repeated the same for GM sT1w/T2w maps.

We performed pairwise spatial cross-correlations between components from the MS and HC cohorts to determine whether similar networks were present also in healthy participants.

Statistical analysis

We used Independent t-test to assess differences between MS and HC in GM and WM maps.

Results

Demographics

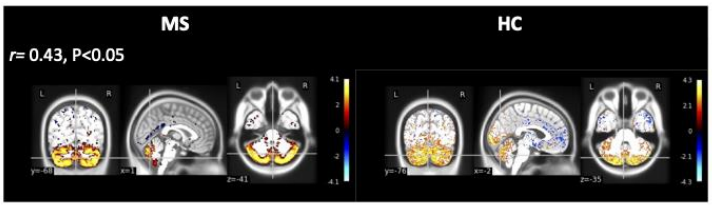
We used data for 300 participants from the Human Connectome Project (HCP). After visually inspecting the results of each MRI processing step, the final sample included 282 participants (mean age 28.8 (SD= 3.5), M= 133, F= 149).

Comparison of WM and GM sT1w/T2w maps and networks in participants with MS and HC

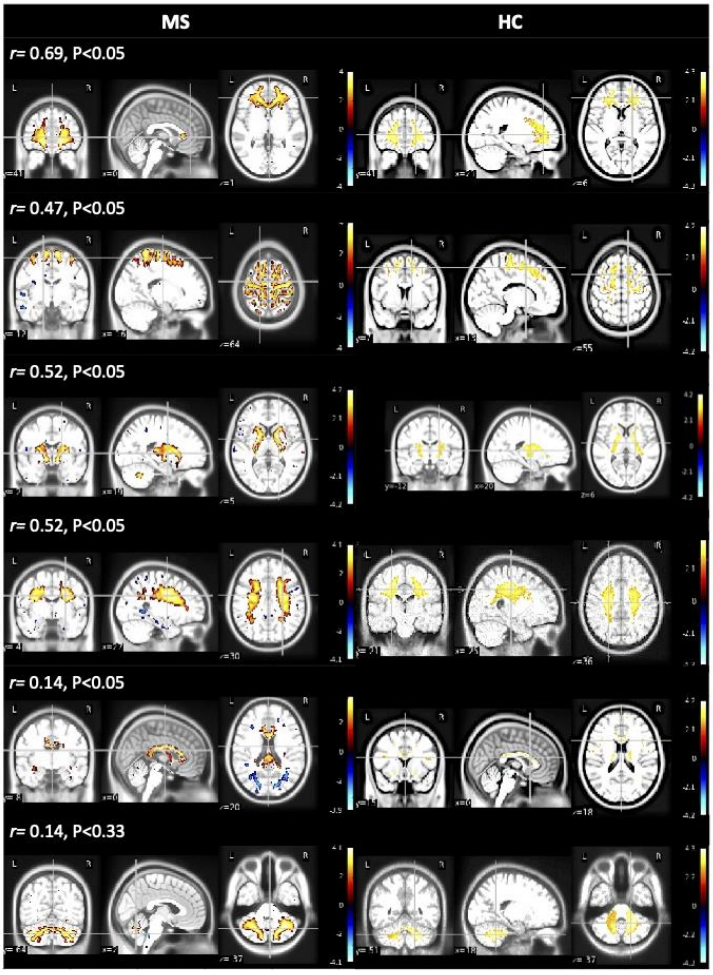
Participants with MS showed a statistically significant decrease in sT1w/T2w measures compared to HC both in GM (respectively, mean (SD)= 0.05(0.01) vs. mean(SD) = 0.12(0.01), P<0.0001) and in WM (respectively, mean(SD)= 0.18(0.05) vs. mean(SD) = 0.32(0.02), P<0.0001) maps (Figure s1).

One GM component (GM-IC4) and six WM networks (WM-IC1, WM-IC2, WM-IC4, WM-IC7, WM-IC8) were similarly identified both in the MS and HC cohorts.

A)


MS
HC
r= 0.43, P<0.05

B)


MS
HC
r= 0.69, P<0.05
r= 0.47, P<0.05
r= 0.52, P<0.05
r= 0.52, P<0.05
r= 0.14, P<0.05
r= 0.14, P<0.33

1

Comparing different calibration techniques

Different standardization techniques have been developed to calibrate the T1/T2 ratio measure. However, so far no gold-standard is available. We aimed to determine the consistency of the standardization technique adopted for this study relative to other standardization techniques.

MRI processing

We randomly selected data for 100 participants and processed MRI images as in Ganzetti et al. (2014) and Margoni et al. (2022). Specifically, we:

1. Manually segmented eye and temporal muscle masks in the MNI template;
2. Registered the MNI template to the halfspace between the T1 and T2 bias-field corrected images using ANTs, and transformed the eye and temporal masks to this space. We visually inspected the registrations and manually corrected the masks were needed.
3. Extracted the distribution peaks (modes) of intensities in the eye (XS) and temporal (YS) masks in the T1w and T2w images of the subject;
4. Extracted the distribution peaks (modes) of intensities in the eye (XR) and temporal (YR) masks in the T1w and T2w images of the MNI
5. Standardized the intensities in T1w and T2w by applying the following formula:

$$IC = \left[\frac{XR-YR}{XS-YS}\right]*I + \left[\frac{XSYR-XRYS}{XS-YS}\right]$$

   where IC is the calibrated image, and I the T1w and T2w images respectively.
6. Divided the calibrated T1 and T2 images to obtain the standardized T1/T2 ratio maps
7. Obtained the standardized T1/T2 ratio maps in GM and WM by applying a GM and WM mask.

We performed ICC analysis between the sT1/T2 ratio measures obtained with this method and the one adopted in this paper to determine the consistency of the measure in GM and WM.

Results

We excluded three participants for whom the estimation of the sT1/T2 ratio was above the 99th percentile. The final sample included 97 participants (mean age: 47.3(SD=9.8); M=32, F=49,

mean duration: 8.5(SD=8.9).The two methods showed an ICC of 0.70 (95% CI[0.57:0.80], p<0.0001) in GM and 0.79 (95% CI[0.68 : 0.86], p<0.0001) in WM.

Estimating confirmed disability progression

Participants did not have treatment change for the entire follow-up.

We estimated the EDSS progression as an increase of 1 point from a baseline EDSS score of 5.5 or below, or of 0.5 points from baseline EDSS score greater than 5.5, and whether the worsening was confirmed at 24 weeks (24-week CDP).[9] We excluded clinical visits within 30 days of an MS relapse, where clinical attack dates where available.

Supplementary Tables

Table s1. Demographics

Discovery Cohort				
Phenotype	*ALL (n= 843)*	*RRMS (n= 159)*	*SPMS (n= 537)*	*PPMS (n= 147)*
Gender (M/F)	*310/533*	*48/111*	*178/359*	*84/63*
Age (years)	*46.6 ± 9.3*	*40.7±10.2*	*48.5±8.4*	*46.4±8.7*
EDSS (median, rage)	*5.5 [0-8]*	*2.5 [0-8]*	*6 [3-7]*	*5 [2-6.5]*
SDMT score (mean, SD)	*42.3±14.7*	*48.0±18.8*	*41.5±13.9*	*NA*
Disease duration (years)	*10.2±8.4*	*7.1±8.9*	*12.9±7.9*	*3.4±3.6*
Number of relapse rate during trial (mean (SD))	*0.51(1.22)*	*0.55(1.63)*	*1.21(2.21)*	*1.7(3.14)*
DMT	*Placebo = 395* *Dimethyl Fumarate 240 mg BID = 20* *Dimethyl Fumarate 240 mg TID = 19* *Lamotrigine= 4* *Glatiramer Acetate = 15* *Interferon beta-1a = 78* *Natalizumab 300 mg = 212* *Fluoxetine = 38* *Amiloride = 33* *Riluzole= 29*	*Placebo = 23* *Dimethyl Fumarate 240 mg BID = 20* *Dimethyl Fumarate 240 mg TID = 19* *Lamotrigine= 4* *Glatiramer Acetate = 15* *Interferon beta-1a = 78*	*Natalizumab 300 mg = 212* *Placebo = 225* *Fluoxetine = 38* *Riluzole= 29* *Amiloride = 33*	*Placebo = 147*
Lesion load (ml)	*28.5.0±25.0*	*23.8.3±24.7*	*31.2±25.5*	*23.7±22.1*
Whole-brain GM volume (ml)	*558.2±75.8*	*559.3±58.8*	*554.9±83.6*	*569.0±59.9*

Replication Cohort				
Phenotype	*ALL (n= 842)*	*RRMS (n=159)*	*SPMS (n=536)*	*PPMS (n=147)*
Gender (M/F)	*329/513*	*45/114*	*220/316*	*64/83*
Age (years)	*46.2± 9.6*	*39.1±10.1*	*48.4±8.4*	*46.1±9.4*
EDSS (median, rage)	*5.5 [0-8]*	*2.5 [0-8]*	*6[3-7.5]*	*4.5[2-6.5]*
SDMT score (mean, SD)	*43.8±14.3*	*48.4±15.9*	*43.3±14.0*	*NA*
Disease duration (years)	*9.7±8.1*	*6.1±8.8*	*12.6±7.2*	*3.0±3.7*
Number of relapse rate during trial (mean(SD))	*0.37(0.67)*	*0.26(0.63)*	*0.25(0.72)*	*0.25(0.57)*

DMT	*Placebo = 416* *Dimethyl Fumarate 240 mg BID = 22* *Dimethyl Fumarate 240 mg TID = 15* *Lamotrigine= 7* *Glatiramer Acetate = 15* *Interferon beta-1a = 78* *Natalizumab 300 mg = 197* *Fluoxetine = 27* *Amiloride = 28* *Riluzole= 28*	*Placebo = 22* *Dimethyl Fumarate 240 mg BID = 22* *Dimethyl Fumarate 240 mg TID = 15* *Lamotrigine= 7* *Glatiramer Acetate = 15* *Interferon beta-1a = 78*	*Natalizumab 300 mg = 197* *Placebo = 247* *Fluoxetine = 27* *Riluzole = 37* *Amiloride = 28*	*Placebo= 147*
Lesion load (ml)	*27.9±27.7*	*19.5±19.5*	*30.9±26.2*	
Whole-brain GM volume (ml)	*560.2±71.7*	*565.1±54.4*	*560.9±78.7*	*552.6±60.6*

Table s2. Description of WM and GM components

GM Network components:	
GM IC1	= anterior cingulate-precuneus-cerebellum pattern
GM IC2	= temporal GM pattern
GM IC3	= fronto-occipital-somatosensory and motor-cerebellar GM pattern
GM IC4	= cerebellum GM pattern
GM IC5	= deep grey matter pattern
GM IC6	= fronto-parieto-cingulate-precuneus-cerebellar pattern
GM IC7	= temporo-parieto-frontal GM pattern
WM Network components:	
WM IC1	= anterior corona radiata regions pattern
WM IC2	= sensorimotor WM pattern
WM IC3	= arcuate fasciculus pattern
WM IC4	= caudate nucleus and internal capsule pattern
WM IC5	= caudate nucleus and cerebellum pattern
WM IC6	= cortical projections pattern
WM IC7	= corpus callosum pattern
WM IC8	= cerebellum WM pattern

Acronyms: GM, grey matter; WM, white matter

Table s3. Correlations between WM and GM networks and GM and WM volume and lesions load

Network	Whole cohort				RRMS				SPMS				PPMS			
	Discovery		Replication		Discovery		Replication		Discovery		Replication		Discovery		Replication	
	Volume	lesion	Volume	lesion	Volume	lesion	Volume	lesion	Volume	lesion	Volume	lesion	Volume	lesion	Volume	lesion
WM 1	r=0.18; [0.11:0.24]; P<0.001*	r=0.39; [0.33:0.44]; P<0.0001*	r=0.09; [0.02:0.16]; P<0.01*	r=0.41; CI0.35:0.46]; P<0.0001*	r=-0.15; [0.3:0.01]; P=0.06	r=-0.33; [0.46:-0.19]; P<0.0001*	r=-0.15; [0.3:0.01]; P=0.058	r=-0.48; [0.59:-0.35]; P<0.0001*	r=0.17; CI0.09:0.25]; P<0.0001*	r=0.4; CI0.33:0.47]; P<0.0001*	r=0.04; [0.05:0.12]; P=0.37	r=0.45; CI0.38:0.52]; P<0.0001*	r=0.12; [0.04:0.28]; P=0.15	r=0.53; CI0.41:0.64]; P<0.0001*	r=0.27; CI0.11:0.42]; P<0.001*	r=0.29; CI0.13:0.43]; P<0.0001*
WM 2	r=0.21; CI0.14:0.27]; P<0.001*	r=0.02; [0.05:0.09]; P=0.60	r=0.14; CI0.07:0.2]; P<0.0001*	r=-0.02; [0.09:0.05]; P=0.57	r=0.15; [0.01:0.3]; P=0.06	r=-0.17; [0.31:-0.01]; P<0.05*	r=-0.02; [0.17:0.14]; P=0.85	r=-0.21; [0.35:-0.06]; P<0.01*	r=0.25; CI0.17:0.32]; P<0.0001*	r=0.08; [0.01:0.16]; P=0.07	r=0.2; CI0.12:0.28]; P<0.0001*	r=0.05; [0.04:0.13]; P=0.29	r=-0.11; [0.27:0.05]; P=0.18	r=-0.17; [0.33:-0.01]; P<0.05*	r=-0.18; [0.34:-0.02]; P=0.05*	r=-0.26; [0.4:-0.1]; P=0.05*
WM 3	r=0.03; [0.04:0.09]; P=0.47	r=-0.71; [0.74:-0.67]; P<0.0001*	r=-0.04; [0.11:0.02]; P=0.20	r=-0.6; [0.64:-0.55]; P<0.0001*	r=0.11; [0.05:0.26]; P=0.19	r=-0.63; [0.72:-0.53]; P<0.0001*	r=0.05; [0.1:0.21]; P=0.50	r=-0.78; [0.83:-0.71]; P<0.0001*	r=-0.02; [0.11:0.06]; P=0.59	r=-0.73; [0.77:-0.69]; P<0.0001*	r=-0.07; [0.15:0.02]; P=0.12	r=-0.65; [0.7:-0.6]; P<0.0001*	r=0.02; [0.14:0.18]; P=0.82	r=-0.66; [0.75:-0.56]; P<0.0001*	r=-0.13; [0.28:0.04]; P=0.13	r=-0.36; [0.49:-0.21]; P<0.0001*
WM 4	r=0.28; CI0.22:0.34]; P<0.0001*	r=-0.5; [0.55:-0.44]; P<0.0001*	r=0.25; CI0.19:0.31]; P<0.0001*	r=-0.49; [0.54:-0.44]; P<0.0001*	r=0.24; CI0.08:0.38]; P<0.005*	r=-0.46; [0.57:-0.33]; P<0.0001*	r=0.18; CI0.02:0.32]; P<0.05*	r=-0.52; [0.63:-0.4]; P<0.0001*	r=0.3; CI0.22:0.37]; P<0.0001*	r=-0.51; [0.57:-0.44]; P<0.0001*	r=0.29; CI0.22:0.37]; P<0.0001*	r=-0.53; [0.59:-0.47]; P<0.0001*	r=-0.01; [0.17:0.15]; P=0.92	r=-0.38; [0.51:-0.23]; P<0.0001*	r=0; [0.16:0.17]; P=0.96	r=-0.33; [0.46:-0.17]; P<0.0001*
WM 5	r=0.35; CI0.29:0.41]; P<0.0001*	r=-0.16; [0.23:-0.09]; P<0.0001*	r=0.31; CI0.25:0.37]; P<0.0001*	r=-0.34; [0.39:-0.27]; P<0.0001*	r=0.32; CI0.17:0.45]; P<0.0001*	r=-0.22; [0.36:-0.06]; P<0.01*	r=0.37; CI0.22:0.49]; P<0.0001*	r=-0.27; [0.41:-0.12]; P<0.001*	r=0.35; CI0.27:0.42]; P<0.0001*	r=-0.12; [0.2:-0.03]; P<0.01*	r=0.31; CI0.23:0.38]; P<0.0001*	r=-0.37; [0.44:-0.29]; P<0.0001*	r=-0.11; [0.26:0.06]; P=0.20	r=-0.01; [0.18:0.15]; P=0.86	r=0.29; CI0.14:0.43]; P<0.0001*	r=-0.19; [0.34:-0.03]; P<0.05*
WM 6	r=0.13; CI0.06:0.19]; P<0.0001*	r=0.43; CI0.37:0.49]; P<0.0001*	r=0.14; CI0.07:0.21]; P<0.0001*	r=0.24; CI0.17:0.3]; P<0.0001*	r=-0.12; [0.27:0.03]; P=0.12	r=-0.4; [0.52:-0.26]; P<0.0001*	r=-0.07; [0.23:0.08]; P=0.35	r=-0.31; [0.45:-0.17]; P<0.0001*	**r=0.12; CI0.04:0.21];** P<0.005*	r=0.48; CI0.41:0.55]; P<0.0001*	r=0.13; CI0.05:0.22]; P<0.005*	r=0.26; CI0.18:0.34]; P<0.0001*	**r=0.14; [0.03:0.29]; P=0.10**	r=0.3; CI0.14:0.44]; P<0.0001*	r=0.2; CI0.04:0.35]; P<0.05*	r=0.27; CI0.11:0.41]; P<0.001*
WM 7	r=0.15; CI0.08:0.21]; P<0.0001*	r=-0.22; [0.28:-0.15]; P<0.0001*	r=0.03; [0.04:0.1]; P=0.39	r=-0.4; [0.45:-0.34]; P<0.0001*	r=0.14; [0.01:0.29]; P=0.07	r=-0.22; [0.37:-0.07]; P<0.005*	r=0.24; CI0.09:0.39]; P<0.005*	r=-0.42; [0.54:-0.29]; P<0.0001*	r=0.12; CI0.03:0.2]; P<0.005*	r=-0.19; [0.28:-0.11]; P<0.0001*	r=-0.02; [0.1:0.07]; P=0.71	r=-0.41; [0.48:-0.34]; P<0.0001*	r=0.29; CI0.13:0.43]; P<0.0001*	r=-0.32; [0.46:-0.16]; P<0.0001*	r=0.1; [0.06:0.26]; P=0.22	r=-0.35; [0.48:-0.2]; P<0.0001*
WM 8	r=0.12; CI0.05:0.19]; P<0.0001*	r=-0.19; [0.26:-0.13]; P<0.0001*	r=0.17; CI0.11:0.24]; P<0.0001*	r=-0.12; [0.19:-0.05]; P<0.0001*	r=0.2; CI0.04:0.34]; P<0.05*	r=-0.12; [0.27:0.04]; P=0.14	r=0.27; CI0.11:0.4]; P<0.0001*	r=-0.05; [0.2:0.11]; P=0.56	**r=0.06; [0.03:0.14]; P=0.20**	r=-0.23; [0.31:-0.15]; P<0.0001*	r=0.11; CI0.03:0.2]; P<0.01*	r=-0.15; [0.23:-0.06]; P<0.01*	r=0.11; [0.05:0.27]; P=0.19	r=0; [0.16:0.16]; P=0.99	r=0.26; CI0.1:0.4]; P<0.05*	r=0; [0.16:0.16]; P=0.99
GM 1	r=0.15; CI0.08:0.21]; P<0.	r=-0.48; [0.53:-0.43]; P<0.0001*	r=0.06; [0.01:0.13]; P=0.08	r=-0.38; [0.44:-0.32]; P<0.0001*	r=0.26; CI0.11:0.4]; P<0.01*	r=-0.47; [0.58:-0.34]; P<0.	r=0.04; [0.12:0.2]; P=0.61	r=-0.47; [0.58:-0.34]; P<0.	r=0.12; CI0.04:0.21]; P<0.005*	r=-0.5; [0.56:-0.43]; P<0.0001*	r=0.06; [0.02:0.15]; P=0.14	r=-0.39; [0.46:-0.31]; P<0.0001*	r=0.16; CI0:0.32]; P=0.052	r=-0.33; [0.46:-0.17]; P<0.0001*	r=0.12; [0.04:0.27]; P=0.15	r=-0.33; [0.46:-0.17]; P<0.0001*

	0001*					0001*		0001*								
GM2	r=0.19; CI0.12:0.25]; P<0.0001*	r=-0.23; [0.29:-0.16]; P<0.00 01*	r=0.14; CI0.08:0.21]; P<0.0001*	r=-0.22; [0.28:-0.15]; P<0.0001*	r=0.31; CI0.16:0.44]; P<0.0001*	r=-0.32; [0.46:-0.18]; P<0.0001*	r=0.18; CI0.02:0.33]; P<0.05*	r=-0.22; [0.36:-0.06]; P<0.01*	r=0.15; CI0.07:0.24]; P<0.0001*	r=-0.21; [0.29:-0.12]; P<0.0001*	r=0.13; CI0.04:0.21]; P<0.005*	r=-0.23; [0.31:-0.15]; P<0.0001*	r=0.24; CI0.08:0.39]; P<0.005*	r=-0.16; [0.31:0]; P=0.057	r=0.26; CI0.1:0.41]; P<0.001*	r=-0.22; [0.36:-0.06]; P<0.01*
GM3	r=0.22; CI0.15:0.28]; P<0.0001*	r=-0.11; [0.18:-0.04]; P<0.005*	r=0.02; [0.05:0.09]; P=0.61	r=-0.12; [0.18:-0.05]; P<0.001*	r=0.09; [0.06:0.25]; P=0.24	r=0.05; [0.11:0.21]; P=0.52	r=-0.06; [0.21:0.1]; P=0.48	r=-0.25; [0.39:-0.1]; P<0.001*	r=0.25; CI0.17:0.33]; P<0.0001*	r=-0.15; [0.23:-0.06]; P<0.001*	r=0.02; [0.06:0.1]; P=0.64	r=-0.1; [0.19:-0.01]; P<0.05*	r=0.17; CI0.01:0.33]; P<0.05*	r=-0.2; [0.35:-0.04]; P=<0.05*	r=0.11; [0.06:0.26]; P=0.20	r=-0.11; [0.27:0.05]; P=0.19
GM4	r=0.04; [0.03:0.1]; P=0.29	r=0.06; [0.01:0.13]; P=0.07	r=-0.05; [0.12:0.02]; P=0.14	r=-0.13; [0.2:-0.06]; P<0.0001*	r=0.1; [0.05:0.26]; P=0.19	r=-0.34; [0.47:-0.2]; P<0.0001*	r=-0.01; [0.16:0.15]; P=0.95	r=-0.39; [0.51:-0.25]; P<0.0001*	r=0.07; [0.02:0.15]; P=0.13	r=-0.01; [0.1:0.07]; P=0.74	r=-0.05; [0.13:0.04]; P=0.29	r=-0.11; [0.2:-0.02]; P<0.05*	r=0.05; [0.11:0.21]; P=0.55	r=-0.01; [0.17:0.16]; P=0.95	r=0.14; [0.02:0.3]; P=0.09	r=-0.02; [0.18:0.14]; P=0.78
GM5	r=0.32; CI0.25:0.38]; P<0.0001*	r=-0.39; [0.45:-P<0.0001*0.33]; P<0.0001*	r=0.32; CI0.25:0.38]; P<0.0001*	r=-0.29; [0.36:-0.23]; P<0.0001*	r=0.08; [0.08:0.23]; P=0.33	r=-0.32; [0.45:-0.18]; P<0.0001*	r=0.17; CI0.01:0.31]; P<0.05*	r=-0.06; [0.21:0.1]; P=0.45	**r=0.37; CI0.3:0.44];** P<0.0001*	r=-0.42; [0.49:-0.34]; P<0.0001*	r=0.34; CI0.27:0.42]; P<0.0001*	r=-0.35; [0.43:-0.27]; P<0.0001*	r=0.12; [0.04:0.28]; P=0.15	r=-0.52; [0.63:-0.39]; P<0.0001*	r=0.24; CI0.08:0.38]; P<0.005*	r=-0.48; [0.6:-0.35]; P<0.0001*
GM6	r=0.12; CI0.05:0.19]; P<0.0001*	r=-0.13; [0.19:-0.06]; P<0.0001*	r=0.06; CI0:0.13]; P=0.07	r=-0.03; [0.09:0.04]; P=0.47	r=0.06; [0.1:0.22]; P=0.44	r=-0.02; [0.18:0.13]; P=0.78	r=0.12; [0.03:0.27]; P=0.12	r=0.1; [0.06:0.25]; P=0.21	r=0.15; CI0.07:0.23]; P<0.0001*	r=-0.17; [0.25:-0.08]; P<0.0001*	r=0.05; [0.04:0.13]; P=0.29	r=-0.06; [0.14:0.03]; P=0.21	r=-0.03; [0.19:0.13]; P=0.71	r=-0.08; [0.24:0.08]; P=0.32	r=0.15; [0.01:0.3]; P=0.07	r=-0.15; [0.3:0.02]; P=0.08
GM7	r=0.15; CI0.08:0.21]; P<0.0001*	r=-0.13; [0.2:-0.06]; P<0.0001*	r=0.16; CI0.1:0.23]; P<0.0001*	r=-0.14; [0.21:-0.08]; P<0.0001*	r=0.17; CI0.01:0.31]; P<0.05*	r=-0.08; [0.23:0.08]; P=0.32	r=0.37; CI0.23:0.5]; P<0.0001*	r=-0.08; [0.23:0.08]; P=0.34	r=0.14; CI0.06:0.23]; P<0.001*	r=-0.1; [0.19:-0.02]; P<0.05*	r=0.14; CI0.06:0.22]; P<0.0001*	r=-0.16; [0.25:-0.08]; P<0.0001*	r=0.09; [0.07:0.25]; P=0.28	r=-0.12; [0.28:0.04]; P=0.15	r=0.14; [0.02:0.3]; P=0.09	r=-0.05; [0.21:0.12]; P=0.58

Table s4. Predicting 24-week CDP using GM and WM components

Network	Whole discovery cohort	Whole replication cohort	RRMS discovery	RRMS replication	SPMS discovery	SPMS replication	PPMS discovery	PPMS replication
WM IC1	R=0.85; 95%CI[0.73:0.99]; P<0.05*	HR=0.84; 95%CI[0.73:0.97]; P<0.05*	HR=0.82; 95%CI[0.53:1.27]; P=0.37	HR=0.55; 95%CI[0.33:0.92]; P<0.05*	HR=0.87; 95%CI[0.72:1.05]; P=0.14	HR=0.83; 95%CI[0.7:0.99]; P<0.05*	HR=0.85; 95%CI[0.6:1.22]; P=0.39	HR=0.92; 95%CI[0.62:1.37]; P=0.69
WM IC2	HR=0.86; 95%CI[0.72:1.03]; P=0.11	HR=0.96; 95%CI[0.82:1.12]; P=0.58	**HR=0.26; 95%CI[0.14:0.48]; P<0.0001***	HR=0.51; 95%CI[0.27:0.97]; P<0.05*	HR=0.89; 95%CI[0.7:1.15]; P=0.38	HR=0.97; 95%CI[0.76:1.25]; P=0.84	HR=0.96; 95%CI[0.47:1.96]; P=0.92	**HR=0.44; 95%CI[0.24:0.82]; P<0.01***
WM IC3	HR=0.88; 95%CI[0.76:1.03]; P=0.12	HR=0.84; 95%CI[0.72:0.97]; P<0.05*	HR=0.52; 95%CI[0.31:0.87]; P<0.05*	HR=0.81; 95%CI[0.48:1.36]; P=0.43	HR=0.95; 95%CI[0.79:1.14]; P=0.58	HR=0.84; 95%CI[0.7:1.01]; P=0.06	HR=0.93; 95%CI[0.59:1.46]; P=0.74	HR=0.88; 95%CI[0.59:1.32]; P=0.53
WM IC6	HR=0.97; 95%CI[0.83:1.14]; P=0.75	HR=0.83; 95%CI[0.71:0.98]; P<0.05*	HR=0.68; 95%CI[0.43:1.09]; P=0.11	HR=0.61; 95%CI[0.35:1.08]; P=0.09	HR=0.97; 95%CI[0.8:1.18]; P=0.76	HR=0.87; 95%CI[0.72:1.05]; P=0.16	HR=0.93; 95%CI[0.58:1.49]; P=0.76	HR=0.76; 95%CI[0.49:1.17]; P=0.21
GM IC1	HR=0.91; 95%CI[0.77:1.08]; P=0.28	HR=0.95; 95%CI[0.79:1.13]; P=0.55	HR=0.46; 95%CI[0.25:0.86]; P<0.05*	HR=0.8; 95%CI[0.4:1.59]; P=0.53	HR=1; 95%CI[0.81:1.23]; P=0.97	HR=0.96; 95%CI[0.78:1.19]; P=0.70	HR=0.87; 95%CI[0.45:1.69]; P=0.69	HR=0.96; 95%CI[0.51:1.81]; P=0.90
GM IC3	HR=0.85; 95%CI[0.7:1.04]; P=0.12	HR=0.87; 95%CI[0.75:1.02]; P=0.08	HR=0.78; 95%CI[0.38:1.6]; P=0.49	HR=0.56; 95%CI[0.24:1.3]; P=0.18	HR=0.73; 95%CI[0.57:0.95]; P<0.05*	HR=0.86; 95%CI[0.72:1.03]; P=0.10	HR=0.68; 95%CI[0.32:1.42]; P=0.30	HR=0.93; 95%CI[0.49:1.75]; P=0.81
GM IC4	HR=0.88; 95%CI[0.75:1.04]; P=0.14	HR=0.98; 95%CI[0.85:1.12]; P=0.75	HR=0.62; 95%CI[0.4:0.96]; P<0.05*	HR=0.92; 95%CI[0.56:1.49]; P=0.73	HR=0.97; 95%CI[0.79:1.19]; P=0.77	HR=0.99; 95%CI[0.84:1.16]; P=0.90	HR=0.9; 95%CI[0.61:1.34]; P=0.61	HR=0.92; 95%CI[0.6:1.41]; P=0.69
GM IC5	HR=0.95; 95%CI[0.79:1.14]; P=0.59	HR=0.83; 95%CI[0.69:1]; P<0.05*	**HR=0.34; 95%CI[0.16:0.71]; P<0.005***	HR=0.42; 95%CI[0.17:1.06]; P=0.07	HR=1; 95%CI[0.8:1.26]; P=0.99	HR=0.83; 95%CI[0.65:1.07]; P=0.16	HR=0.83; 95%CI[0.42:1.63]; P=0.59	HR=0.98; 95%CI[0.42:2.26]; P=0.96
GM IC7	HR=0.83; 95%CI[0.7:0.98]; P<0.05*	**HR=0.8; 95%CI[0.69:0.92]; P<0.005***	HR=0.77; 95%CI[0.47:1.26]; P=0.31	HR=0.69; 95%CI[0.41:1.15]; P=0.15	HR=0.91; 95%CI[0.74:1.12]; P=0.38	**HR=0.76; 95%CI[0.64:0.91]; P<0.005***	**HR=0.51; 95%CI[0.32:0.81]; P<0.005***	HR=0.95; 95%CI[0.65:1.37]; P=0.77
Lesion load	HR=1.12; 95%CI[0.96:1.3]; P=0.14	HR=1.02; 95%CI[0.88:1.18]; P=0.79	HR=1.4; 95%CI[1:1.95]; P<0.05*	HR=1.45; 95%CI[0.87:2.41]; P=0.15	HR=1.03; 95%CI[0.85:1.25]; P=0.77	HR=1.01; 95%CI[0.82:1.24]; P=0.92	HR=1.02; 95%CI[0.7:1.47]; P=0.93	HR=1.07; 95%CI[0.82:1.4]; P=0.62
Whole brain GM volume	HR=0.93; 95%CI[0.8:1.09]; P=0.37	HR=0.94; 95%CI[0.78:1.13]; P=0.50	HR=0.98; 95%CI[0.55:1.76]; P=0.96	HR=0.7; 95%CI[0.39:1.24]; P=0.22	HR=0.92; 95%CI[0.77:1.1]; P=0.36	HR=0.92; 95%CI[0.73:1.17]; P=0.5	HR=0.93; 95%CI[0.56:1.55]; P=0.78	HR=0.75; 95%CI[0.48:1.15]; P=0.19
Deep GM volume	HR=0.99; 95%CI[0.83:1.17]; P=0.89	HR=0.98; 95%CI[0.83:1.14]; P=0.76	HR=0.85; 95%CI[0.49:1.46]; P=0.55	HR=0.94; 95%CI[0.64:1.39]; P=0.77	HR=0.95; 95%CI[0.75:1.18]; P=0.63	HR=0.84; 95%CI[0.66:1.07]; P=0.15	HR=0.77; 95%CI[0.53:1.12]; P=0.17	HR=0.7; 95%CI[0.45:1.09]; P=0.12
Thalamus	HR=0.95; 95%CI[0.8:1.13]; P=0.55	HR=0.98; 95%CI[0.83:1.16]; P=0.85	HR=0.87; 95%CI[0.55:1.38]; P=0.56	HR=0.86; 95%CI[0.53:1.38]; P=0.52	HR=0.99; 95%CI[0.81:1.21]; P=0.92	HR=0.94; 95%CI[0.75:1.16]; P=0.55	HR=0.87; 95%CI[0.54:1.4]; P=0.57	HR=0.75; 95%CI[0.48:1.17]; P=0.20
Caudate	HR=0.92; 95%CI[0.77:1.09]; P=0.33	HR=0.9; 95%CI[0.76:1.07]; P=0.24	HR=0.67; 95%CI[0.41:1.08]; P=0.10	HR=0.84; 95%CI[0.52:1.36]; P=0.48	HR=0.99; 95%CI[0.81:1.21]; P=0.93	HR=0.87; 95%CI[0.69:1.09]; P=0.22	HR=0.9; 95%CI[0.58:1.41]; P=0.65	HR=0.83; 95%CI[0.54:1.27]; P=0.40
Pallidum	HR=0.96; 95%CI[0.81:1.13]; P=0.64	HR=0.92; 95%CI[0.76:1.1]; P=0.35	HR=0.85; 95%CI[0.52:1.41]; P=0.54	HR=0.86; 95%CI[0.49:1.49]; P=0.58	HR=0.94; 95%CI[0.78:1.13]; P=0.53	HR=0.83; 95%CI[0.65:1.07]; P=0.15	HR=0.88; 95%CI[0.56:1.37]; P=0.57	HR=0.79; 95%CI[0.51:1.22]; P=0.28

***statistically significant**

Table sT2 shows the hazard ratio, 95% CI, and p-value for the 24-week CDP for GM and WM components and lesion load in the discovery and application cohorts' whole, RRMS, SPMS, and PPMS groups. In bold, components that were still statistically significant after correcting for multiple comparisons.

Acronyms: CI, confidence interval, GM, grey matter; WM, white matter

Table s5. Predicting SDMT worsening using GM and WM components

Network	Whole discovery cohort	Whole application cohort	RRMS discovery	RRMS application	SPMS discovery	SPMS application
WM IC1	**HR=0.85; 95%CI[0.75:0.96]; p<0.01***	HR=0.93; 95%CI[0.82:1.04]; p=0.21	HR=0.93; 95%CI[0.56:1.55]; p=0.79	HR=0.73; 95%CI[0.45:1.19]; p=0.21	HR=0.85; 95%CI[0.74:0.98]; p<0.05*	HR=0.94; 95%CI[0.82:1.07]; p=0.33
WM IC2	**HR=0.74; 95%CI[0.64:0.86]; p<0.0001***	HR=0.90; 95%CI[0.78:1.02]; p=0.10	HR=0.85; 95%CI[0.44:1.68]; p=0.65	HR=0.98; 95%CI[0.58:1.66]; p=0.94	HR=0.84; 95%CI[0.69:1.02]; p=0.08	HR=0.96; 95%CI[0.8:1.15]; p=0.64
WM IC3	**HR=0.82; 95%CI[0.72:0.95]; p<0.01***	**HR=0.82; 95%CI[0.72:0.94]; p<0.005***	HR=0.83; 95%CI[0.5:1.39]; p=0.48	HR=0.66; 95%CI[0.39:1.1]; p=0.11	HR=0.82; 95%CI[0.7:0.96]; p<0.05*	HR=0.84; 95%CI[0.73:0.96]; p<0.05*
WM IC4	HR=0.93; 95%CI[0.82:1.05]; p=0.25	HR=0.90; 95%CI[0.8:1.02]; p=0.09	HR=0.89; 95%CI[0.57:1.4]; p=0.62	HR=0.59; 95%CI[0.34:1.01]; p=0.054	**HR=0.74; 95%CI[0.63:0.86]; p<0.0001***	HR=0.84; 95%CI[0.74:0.96]; p<0.05*
WM IC5	HR=1.01; 95%CI[0.89:1.14]; p=0.90	HR=0.90; 95%CI[0.8:1.01]; p=0.08	HR=0.99; 95%CI[0.56:1.78]; p=0.98	HR=0.45; 95%CI[0.23:0.88]; p<0.05*	**HR=0.79; 95%CI[0.68:0.92]; p<0.005***	**HR=0.8; 95%CI[0.69:0.93]; p<0.005***
WM IC7	HR=0.87; 95%CI[0.76:0.99]; p<0.05*	**HR=0.73; 95%CI[0.64:0.84]; p<0.0001***	HR=0.73; 95%CI[0.43:1.25]; p=0.26	HR=0.65; 95%CI[0.38:1.11]; p=0.12	HR=0.92; 95%CI[0.79:1.08]; p=0.31	**HR=0.75; 95%CI[0.65:0.88]; p<0.0001***
GM IC1	HR=0.99; 95%CI[0.87:1.12]; p=0.86	HR=0.96; 95%CI[0.85:1.09]; p=0.53	HR=0.94; 95%CI[0.47:1.86]; p=0.86	HR=0.36; 95%CI[0.15:0.87]; p<0.05*	HR=0.84; 95%CI[0.72:0.99]; p<0.05*	HR=0.89; 95%CI[0.77:1.03]; p=0.13
GM IC2	HR=0.97; 95%CI[0.86:1.1]; p=0.62	HR=0.87; 95%CI[0.78:0.98]; p<0.05*	HR=0.61; 95%CI[0.37:0.99]; p<0.05*	HR=0.83; 95%CI[0.5:1.37]; p=0.47	HR=0.99; 95%CI[0.86:1.13]; p=0.87	HR=0.88; 95%CI[0.77:1]; p<0.05*
GM IC3	**HR=0.72; 95%CI[0.62:0.84]; p<0.0001***	HR=0.97; 95%CI[0.87:1.09]; p=0.63	HR=0.72; 95%CI[0.33:1.55]; p=0.40	HR=0.89; 95%CI[0.44:1.83]; p=0.76	HR=0.79; 95%CI[0.65:0.97]; p<0.05*	HR=0.98; 95%CI[0.86:1.11]; p=0.70
GM IC5	**HR=0.75; 95%CI[0.64:0.88]; p<0.001***	**HR=0.80; 95%CI[0.69:0.93]; p<0.005***	HR=0.78; 95%CI[0.38:1.6]; p=0.50	HR=0.9; 95%CI[0.46:1.76]; p=0.76	HR=0.83; 95%CI[0.68:1.02]; p=0.08	HR=0.81; 95%CI[0.67:0.99]; p<0.05*
GM IC6	HR=0.86; 95%CI[0.75:1]; p<0.05*	**HR=0.85; 95%CI[0.75:0.96]; p<0.01***	HR=0.91; 95%CI[0.36:2.32]; p=0.85	HR=0.71; 95%CI[0.39:1.28]; p=0.26	HR=0.9; 95%CI[0.74:1.09]; p=0.29	HR=0.88; 95%CI[0.77:1]; p=0.057
Whole brain GM volume	HR=0.9; 95%CI[0.78:1.04]; p=0.14	HR=0.88; 95%CI[0.76:1.02]; p=0.10	HR=0.92; 95%CI[0.53:1.62]; p=0.78	HR=0.9; 95%CI[0.49:1.66]; p=0.74	HR=0.93; 95%CI[0.79:1.09]; p=0.38	HR=0.91; 95%CI[0.77:1.07]; p=0.25
DGM volume	HR=0.97; 95%CI[0.85:1.1]; p=0.63	HR=0.95; 95%CI[0.83:1.08]; p=0.41	HR=0.87; 95%CI[0.47:1.64]; p=0.67	HR=0.83; 95%CI[0.43:1.61]; p=0.58	HR=0.99; 95%CI[0.85:1.14]; p=0.84	HR=0.96; 95%CI[0.83:1.1]; p=0.53
Lesion load	**HR=1.27; 95%CI[1.11:1.44]; p<0.0001***	**HR=1.26; 95%CI[1.11:1.42]; p<0.0001***	HR=1.1; 95%CI[0.74:1.64]; p=0.63	HR=1.46; 95%CI[0.8:2.67]; p=0.22	**HR=1.44; 95%CI[1.24:1.67]; p<0.0001***	**HR=1.27; 95%CI[1.12:1.43]; p<0.0001***
Thalamus	HR=0.87; 95%CI[0.75:1.01]; p=0.07	**HR=0.77; 95%CI[0.66:0.9]; p<0.001***	HR=0.93; 95%CI[0.59:1.49]; p=0.78	HR=0.63; 95%CI[0.39:1.03]; p=0.06	HR=0.79; 95%CI[0.66:0.95]; p<0.05*	**HR=0.77; 95%CI[0.65:0.92]; p<0.005***
Caudate	HR=0.95; 95%CI[0.83:1.08]; p=0.43	HR=0.89; 95%CI[0.77:1.03]; p=0.11	HR=0.88; 95%CI[0.58:1.35]; p=0.56	HR=0.74; 95%CI[0.44:1.23]; p=0.24	HR=0.95; 95%CI[0.82:1.11]; p=0.53	HR=0.93; 95%CI[0.8:1.08]; p=0.36
Pallidum	HR=0.91; 95%CI[0.79:1.06]; p=0.22	HR=0.81; 95%CI[0.68:0.96]; p<0.05*	HR=0.9; 95%CI[0.53:1.53]; p=0.70	HR=0.6; 95%CI[0.35:1.04]; p=0.07	HR=0.87; 95%CI[0.73:1.03]; p=0.11	HR=0.83; 95%CI[0.69:1]; p<0.05*

***statistically significant**

Table sT3 shows the hazard ratio, 95% CI, and p-value for 10% SDMT worsening for GM and WM components, whole brain GM volume loss, and lesion load in the discovery and application cohorts' whole, RRMS, SPMS, and PPMS groups. In bold, components that were still statistically significant after correcting for multiple comparisons.

Acronyms: CI, confidence interval, GM, grey matter; WM, white matter

Supplementary Figures

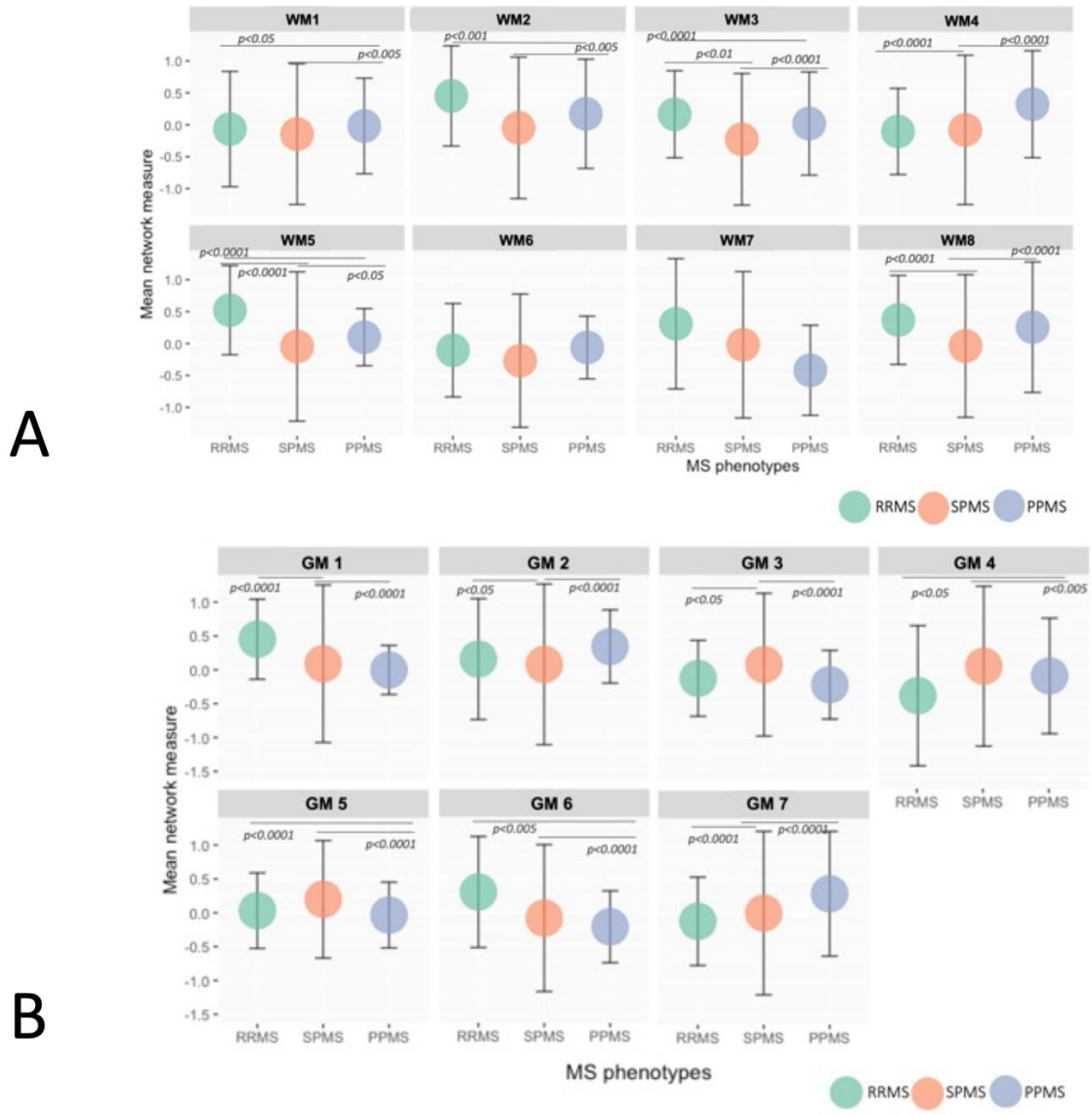
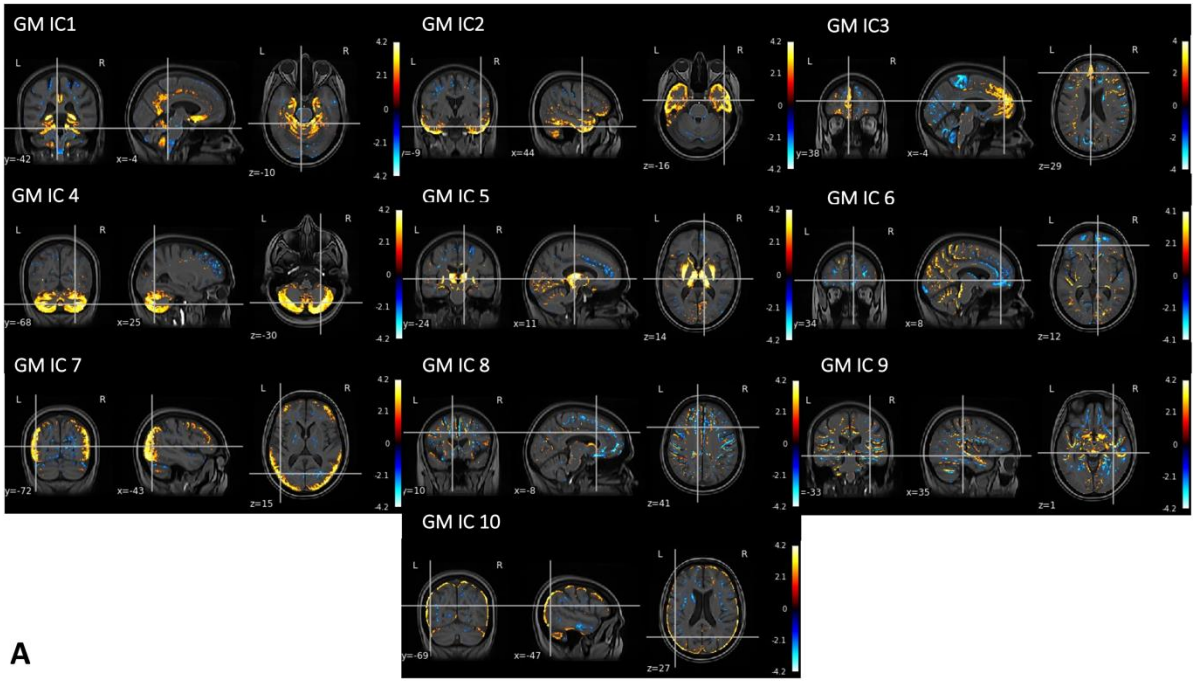


Figure s1. Comparison of WM and GM sT1w/T2w networks across clinical MS phenotypes



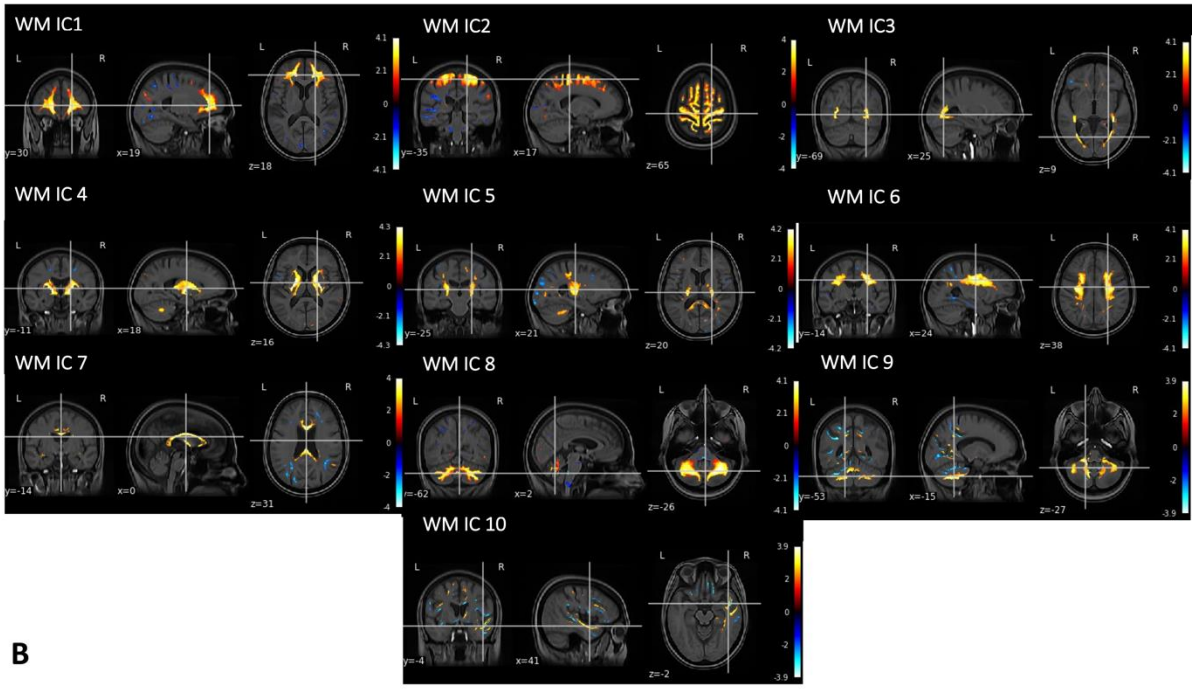


Figure s2. Figure s2 shows stable (**A**) GM and (**B**) WM patterns of covarying sT1w/T2w changes identified by the ICA. The colour bar represents the loading of each network: the higher the loading, the lower the microstructural damage in the network.