What Predicts Cognitive Decline in Multiple Sclerosis?

A 5-year Follow-up Study

A.J.C. Eijlers[1]*, Q. van Geest[1]*, I. Dekker[2,3], M.D. Steenwijk[1], K.A. Meijer[1], H.E. Hulst[1], F. Barkhof[2,4] B.M.J. Uitdehaag[3], M.M. Schoonheim[1] and J.J.G. Geurts[1]

Departments of Anatomy & Neurosciences[1], Radiology and Nuclear Medicine[2], Neurology[3], Amsterdam Neuroscience, MS Center Amsterdam, VU University Medical Center, Amsterdam, the Netherlands. Institutes of Neurology and Healthcare Engineering, UCL, London, UK[4].

* Both authors contributed equally to this article

Corresponding author: Anand Eijlers

Address: VU University Medical Center, O2 building, 13 W01, De Boelelaan 1108, 1081 HZ Amsterdam.

T: +31619987909

E: a.eijlers@vumc.nl

Brief running title: Predicting cognitive decline in MS

Keywords: multiple sclerosis, MRI, cognition, atrophy, longitudinal

Abbreviations

FSL: Functional Magnetic Resonance Imaging of the Brain Software Library

EDSS: Expanded Disability Status Scale

MNI: Montreal Neurological Institute

Abstract

Cognitive decline is common in multiple sclerosis and strongly affects overall quality of life. Despite the identification of cross-sectional MRI correlates of cognitive impairment, predictors of *future* cognitive decline remain unclear. The objective of this study was to identify which MRI measures of structural damage, demographic and/or clinical measures at baseline best predict cognitive decline, during a five year follow-up period. A total of 234 patients with clinically definite multiple sclerosis and 60 healthy controls were examined twice, with a five year interval (mean = 4.9 years, SD = 0.9). An extensive neuropsychological evaluation was performed at both time points and the reliable change index was computed to evaluate cognitive decline. Both whole-brain and regional MRI (3-Tesla) measures were assessed at baseline, including white matter lesion volume, diffusion-based white matter integrity, cortical and deep grey matter volume. Logistic regression analyses were performed to determine which baseline measures best predicted cognitive decline in the entire sample as well as in early relapsing-remitting (symptom duration <10 years), late relapsing-remitting (symptom duration ≥10 years) and progressive phenotypes. At baseline, patients with multiple sclerosis had a mean disease duration of 14.8 (SD=8.4) years and 96/234 patients (41%) were classified as cognitively impaired. A total of 66/234 patients (28%) demonstrated cognitive decline during follow-up, with higher frequencies in progressive compared to relapsing-remitting patients: 18/33 secondary progressive patients (55%), 10/19 primary progressive patients (53%) and 38/182 relapsing-remitting patients (21%). A prediction model that included only *whole-brain* MRI measures (Nagelkerke R^2 = 0.22, P<0.001) showed cortical grey matter volume as the only significant MRI predictor of cognitive decline, while a prediction model that assessed *regional* MRI measures (Nagelkerke R^2 = 0.35, P<0.001) indicated integrity loss of the anterior thalamic radiation, lesions in the superior longitudinal fasciculus and temporal atrophy as significant MRI predictors for cognitive decline. Disease stage specific regressions showed that cognitive decline in early relapsing-remitting multiple sclerosis was predicted by white matter integrity damage, while cognitive decline in late relapsing-remitting and progressive multiple sclerosis was predicted by cortical atrophy. These results indicate that patients with more severe structural damage at baseline, and especially cortical atrophy, are more prone to suffer from cognitive decline. New studies now need to further elucidate the underlying mechanisms leading to cortical atrophy, evaluate the value of including cortical atrophy as a possible outcome marker in clinical trials as well as study its potential use in individual patient management.

Introduction

Multiple sclerosis is an inflammatory, demyelinating and neurodegenerative disease of the central nervous system. It affects an estimated 2.5 million people worldwide, and is the primary cause of disability in young adults in developed countries (Dutta and Trapp, 2011; Dendrou *et al.*, 2015). Besides physical disability, cognitive deficits are present in 40 to 70 percent of patients with multiple sclerosis and have a severe impact on daily functioning (Chiaravalloti and DeLuca, 2008; Benedict *et al.*, 2017b). The most commonly affected cognitive domains include information processing speed and episodic memory, with impairments in executive function, verbal fluency and visuospatial memory also commonly detected (Sumowski *et al.*, 2018). Multiple sclerosis pathology in the form of focal lesions, normal-appearing tissue damage and atrophy affects both the white and grey matter tissue compartments and can be visualized using MRI (Wattjes *et al.*, 2015). While white matter lesions and diffuse damage as well as deep grey matter atrophy already occur early in the disease, cortical demyelination and atrophy seem to predominate in later stages (Bergsland *et al.*, 2012; Haider *et al.*, 2014; Schoonheim *et al.*, 2014; Steenwijk *et al.*, 2014).

How this apparent ordering of types of damage in different disease stages contributes to cognitive decline still remains incompletely understood. White matter lesions have been shown to only weakly to moderately relate to the severity of cognitive deficits, while stronger correlations have been observed for normal appearing white matter damage and grey matter atrophy (Benedict and Zivadinov, 2011; Schoonheim *et al.*, 2014). In addition to the relevance of measuring aggregate damage to white and grey matter tissue compartments to unravel substrates of cognitive impairment, a number of studies has evaluated the significance of damage to specific, cognitively relevant, brain regions or connections. These (mainly cross sectional) studies highlighted thalamic (Minagar *et al.*, 2013), hippocampal (Damjanovic *et al.*, 2017) and cortical regions such as the posterior cingulate cortex as particularly relevant for cognitive decline (Steenwijk *et al.*, 2016). Longitudinal studies investigating cognitive functioning in multiple sclerosis remain scarce, but have shown that cognitive decline is concomitant with progressing lesion volumes (Camp *et al.*, 2005) and whole-brain (Zivadinov *et al.*, 2001) grey and white matter atrophy (Rocca *et al.*, 2018). There still is an urgent clinical need, however, to be able to predict more accurately whether a patient will progress based on baseline measurements, in order to optimize disease management and treatment strategies.

Only a few studies investigated such predictors of cognitive decline in multiple sclerosis, suggesting predictive power of early changes in lesions (Summers *et al.*, 2008b), whole brain atrophy (Summers *et al.*, 2008a; Deloire *et al.*, 2011) and tissue integrity (Deloire *et al.*, 2011; Filippi *et al.*, 2013), all studying relatively early multiple sclerosis cohorts, small sample sizes, 1.5 Tesla scanners and no regional MRI measures. Therefore, in the present study, we first retrospectively assessed whether patients demonstrated a decline in cognitive functioning during a five-year follow-up and subsequently compared baseline characteristics (i.e., MRI, demographic and clinical measures) between cognitively stable and declining patients. Secondly, we investigated which baseline measures best *predicted* future cognitive decline, with separate models evaluating whole-brain and regional MRI predictors. Finally, cross-sectional and longitudinal predictors of cognitive impairment and decline were compared, in order to disentangle predictive and correlative markers. These questions were addressed in a large cohort of 234 patients with relapsing-remitting, secondary progressive and primary progressive multiple sclerosis, as well as 60 healthy controls, who received extensive neuropsychological evaluations and advanced 3 Tesla MRI at baseline and identical neuropsychological testing after five years.

Materials and methods

Participants

For this prospective study, a total of 332 patients with clinically definite multiple sclerosis (Polman *et al.*, 2011), part of the Amsterdam Multiple Sclerosis Cohort (Eijlers *et al.*, 2017; Meijer *et al.*, 2017), and 96 healthy controls received MRI and cognitive evaluation at baseline and 234/332 patients (70%) and 60/96 healthy controls (63%) returned for an identical cognitive evaluation at follow-up. Only the 236 patients with multiple sclerosis (32% men, age 47.61 ± 11.02 years, symptom duration 14.6 ± 8.4 years) and 60 healthy controls (42% men, age 46.45 ± 9.91 years) that received a cognitive follow-up were retained for the current study. The average interval time between baseline and follow-up visits was 4.80 years (SD=0.85) for patients with multiple sclerosis and 5.46 years (SD=1.08) for healthy controls. Clinical phenotypes at baseline included 181 patients with relapsing-remitting, 33 with secondary progressive and 20 with primary progressive multiple sclerosis. Disease modifying treatments at baseline included β-interferons (n=57), glatiramer acetate (n=13), natalizumab (n=12) or other immunosuppressive therapy (n=5). The highest level of education attained was assessed using a scale between one, did not finish primary school, and seven, acquired university degree or higher. Overall disability of patients with multiple sclerosis was assessed using the Expanded Disability Status Scale (Kurtzke, 1983). Patients were relapse-free and without steroid treatment for at least two months prior to both baseline and follow-up visits. The study was approved by the institutional ethics review board of the VU University Medical Center and participants gave written informed consent prior to participation.

Neuropsychological evaluation

Participants underwent extensive neuropsychological evaluation at both time points using an expanded Brief Repeatable Battery of Neuropsychological tests (Rao, 1990) as previously described (Schoonheim *et al.*, 2015). Executive functioning was assessed using the concept shifting test (Van der Elst *et al.*, 2006), with the ascending number ordering, alphabetical letter ordering and alternating letter and number ordering conditions corrected for motor speed, converted into Z-scores and then averaged to create a domain Z-score.. Verbal memory was assessed using the selective reminding test (Buschke, 1973), with the average long-term storage of the first trial, total long-term recall and delayed recall scores converted into Z-scores and averaged. Verbal fluency was assessed using the word list generation test (Boringa *et al.*, 2001), with the total number of correct responses in 60 seconds converted into a Z-

score. Information processing speed was assessed using the symbol digit modalities test (Smith, 1982), with the total number of correctly substituted symbols in 90 seconds converted into a Z-score. Visuospatial memory was assessed using the spatial recall test (Boringa *et al.*, 2001), with the total score on three immediate recall trials and the score on the delayed recall trial converted into Z-scores and averaged. Attention was assessed using the Stroop colour-word test (Stroop, 1992), with the time to complete the first, second and third trial as well as the time to complete the third trial corrected for the time to complete the first and second trials converted into Z-scores and averaged. Finally, working memory was assessed using the memory comparison test (Brand and Jolles, 1987), with the time taken to complete the percent sign, one, two, three, and four letter trials converted into Z-scores and averaged. The cognitive scores of all subjects were corrected for effects of sex, age and education observed in the healthy controls, using a previously published method (Amato *et al.*, 2006). Patients were classified as cognitively impaired if performance was below $Z < -1.5$ on two or more cognitive domains, grouping together mildly and severely impaired patients (Eijlers *et al.*, 2017).

Cognitive change during follow-up: reliable change index

To assess cognitive change in the patients with multiple sclerosis during the follow-up period, the modified practice adjusted reliable change index (Iverson, 2001) was computed, which corrected for practice effects as observed in the healthy control group. Next, reliable change index scores for each cognitive domain were divided by each individual subject's time interval between baseline and follow-up, obtaining a *yearly rate of change on each cognitive domain*. These rates of change were then averaged across domains to obtain a *averaged yearly rate of cognitive change* for each patient. In order to separate patients into *cognitively stable* and *cognitively declining* groups, two different approaches were explored. The first approach was similar to the commonly used criterion to classify patients as cognitively impaired in cross sectional studies, and was based on a decline on at least two separate cognitive domains (Louapre *et al.*, 2014; Schoonheim *et al.*, 2015). The second approach was based on another commonly used measure in cross-sectional studies, namely applying a threshold to average cognitive functioning (using the abovementioned average rate of cognitive change). The final approach to define 'cognitive decline' was chosen based on the most optimal combination of having high sensitivity to detect cognitively declining patients (i.e., a minimal residual decline in the remaining cognitively stable patients with a yearly rate of cognitive change <0.01) and

high specificity (i.e., a low number of healthy controls incorrectly classified as cognitively declining). The different approaches were explored with incremental steps in the rate of cognitive change of 0.05 and the best performing criterion was selected and used to classify patients as 'cognitively declining' for further analyses.

Magnetic resonance imaging

All subjects were scanned on a 3 Tesla whole-body magnetic resonance system (General Electric Signa-HDxt, Milwaukee, WI, USA), using an eight-channel phased-array head coil. The protocol included a three-dimensional T1-weighted fast spoiled gradient echo sequence for volumetric measurements (repetition time 7.8 ms, echo time 3 ms, inversion time 450 ms, flip angle 12 degrees, 1.0 mm sagittal slices, 0.9 x 0.9 mm^2 in-plane resolution), a three-dimensional fluid attenuated inversion recovery sequence for white matter lesion segmentation (repetition time 8000 ms, echo time 125 ms, inversion time 2350 ms, 1.2 mm sagittal slices, 0.98 x 0.98 mm^2 in-plane resolution) and a diffusion tensor imaging sequence for white matter integrity assessment, covering the entire brain (five volumes without directional weighting, i.e. b0, and 30 volumes with non-collinear diffusion gradients, echo planar imaging, b=1000s/mm2, repetition time 13000ms, echo time 91ms, flip angle 90 degrees, 53 contiguous axial slices of 2.4mm, in-plane resolution 2x2mm).

Whole-brain, regional and voxel-wise measures of damage

White matter lesions

White matter lesions were automatically segmented on the fluid attenuated inversion recovery images using k-nearest neighbour classification with tissue type priors (Steenwijk *et al.*, 2013) and used to compute whole-brain lesion volume. *Regional* lesion volumes were defined as the percentage of white matter within individual tracts affected by lesions, which was computed within each of the ten tract masks part of the JHU-ICBM tracts atlas (part of FSL 5) after non-linear registration of individual lesion maps to Montreal Neurological Institute (MNI)-152 standard space. Finally, to detect voxel-wise differences in lesion location between cognitively stable and declining groups, individual lesion maps were smoothed with a 5mm Gaussian filter, followed by a voxel-wise group comparison using RANDOMISE (part of FSL 5)(Winkler *et al.*, 2014), with the analysis constrained to areas with a 5% lesion probability for the entire group (Dalton *et al.*, 2012; Filli *et al.*, 2012).

White matter integrity

Diffusion weighted data were pre-processed using FSL 5, including motion- and eddy current correction on images and gradient-vectors followed by diffusion tensor fitting. To assess white matter integrity, fractional anisotropy maps were computed and non-linearly registered to the FMRIB58_FA brain. Next, fractional anisotropy maps were averaged across subjects and skeletonized to obtain the main white matter tracts common to the group using the standard Tract-Based Spatial Statistics pipeline (part of FSL 5)(Smith *et al.*, 2006). Subsequently, individual subject fractional anisotropy values were projected onto this skeleton and the mean skeleton fractional anisotropy was computed for each individual as a measure of whole-brain white matter integrity. To assess *regional* white matter integrity damage, the integrity of individual white matter tracts was computed for each subject, based on the same approach as for lesions. The ten tract masks part of the JHU-ICBM tracts atlas were overlaid on the individual fractional anisotropy skeletons and the mean integrity within each mask was computed. Z-scores representing effect sizes of damage were then computed for each tract in every patient relative to the healthy control group. Finally, voxel-wise fractional anisotropy values within the white matter skeleton were compared between cognitively stable and declining patients using RANDOMISE.

Deep grey matter atrophy

Before tissue segmentation, white matter lesion masks were registered to the three-dimensional T1-weighted images to enable lesion filling using LEsion Automated Pre-processing (Chard *et al.*, 2010). Whole-brain, grey matter and white matter volumes were calculated on the lesion-filled images using SIENAX (Smith *et al.*, 2002), following the previously published pipeline (Popescu *et al.*, 2012) and deep grey matter volumes were obtained using FIRST (Patenaude *et al.*, 2011)(both part of FSL 5). All volumes were normalised for head size using the V-scaling factor derived by SIENAX. Individual deep grey matter volumes were summed to obtain a whole-brain deep grey matter volume. To assess atrophy of individual deep grey matter regions, left and right volumes were summed and Z-scores reflecting the atrophy effect sizes were computed for each region and each patient relative to the healthy control group. Finally, the shape of individual deep grey matter regions was assessed using the FIRST vertex analysis pipeline (part of FSL 5) in MNI152 standard space and compared between cognitively stable and declining patients using RANDOMISE.

Cortical grey matter atrophy

To calculate whole-brain cortical grey matter volume, individual FIRST deep grey matter segmentation images were subtracted from the SIENAX grey matter segmentation images using fslmaths. To assess regional atrophy of individual cortical lobes, the MNI structural atlas (part of FSL 5) was first nonlinearly registered to the lesion filled three-dimensional T1-weighted images in subject space. Next, lobar structural atlas masks were overlaid on the grey matter segmentation images from SIENAX to compute lobar grey matter volumes (left and right volumes were summed). Z-scores reflecting the effect sizes of regional atrophy were then computed for each cortical lobe in each patient relative to the healthy control group. Finally, voxel-wise cortical grey matter density was compared between cognitively stable and declining patients using the standard voxel-based morphometry pipeline (part of FSL 5)(Good *et al.*, 2001). This includes non-linear registration of grey matter images to MNI152 standard space, multiplication by the Jacobian of the warp field to correct for this non-linear deformation (Douaud *et al.*, 2007) and smoothing with a 4 mm Gaussian filter followed by a voxel-wise group comparison using RANDOMISE.

Statistical analysis

Statistical analyses of the demographic, clinical and whole-brain MRI variables were performed in SPSS version 22 (Armonk, NY, USA). All demographic, clinical and volumetric MRI variables were checked for normality using the Kolmogorov-Smirnov test and histogram inspection; lesion volumes were log-transformed. Nonparametric testing was used to assess group differences for not normally distributed demographic variables and EDSS. A Wilcoxon signed rank test was used to assess longitudinal changes in EDSS and a two-tailed one-sample t-test to assess whether the average rate of cognitive change in patients with multiple sclerosis was significantly different from zero. Multivariate general linear model analyses were performed to assess group differences in normally distributed demographic, whole-brain MRI and cognitive variables, with sex, age and education entered as covariates. Group comparison *P*-values < 0.05 were considered statistically significant after Bonferroni correction for multiple comparisons; corrected p-values were reported. All voxel-wise group comparisons were performed using RANDOMISE (part of FSL 5) with 5000 permutations and included sex, age and education as covariates. Here, multiple comparisons correction was performed using threshold free cluster enhancement and family-wise error correction, with statistical significance threshold of $P<0.05$.

To investigate which baseline measures relate to cross sectional cognitive impairment as well as predict subsequent longitudinal cognitive decline, logistic regression analyses with conditional backward selection were performed using baseline measures only. In the first two regression analyses, the correlative and predictive value of *whole-brain* MRI measures for cross sectional cognitive impairment and longitudinal cognitive decline were evaluated. Additionally, the effects of disease phenotype were investigated by separately running the aforementioned model in early relapsing-remitting multiple sclerosis (symptom duration <10 years, n = 92), late relapsing-remitting multiple sclerosis (symptom duration ≥10 years, n = 90) and progressive multiple sclerosis (pooling primary and secondary progressive phenotypes, n = 52). This analysis was performed for global cognitive decline as described above, but also for information processing speed decline, given the clear clinical importance of this domain in multiple sclerosis. Decline of this domain was based on the recently established criterion for clinically meaningful decline, i.e. a loss of four points on the symbol digit modalities test (Benedict *et al.*, 2017a). Next, the predictive value of *regional* MRI measures for longitudinal global cognitive decline was evaluated in four separate regression analyses; a model that included individual cortical lobar volumes, deep grey matter volumes, white matter tract integrities and white matter tract lesion percentages, after which significant regional MRI predictors of these four models were included in a final combined regional model. Initially, all models included the same clinical and demographic variables, namely: symptom duration, multiple sclerosis phenotype (relapse onset versus primary progressive and relapsing-remitting versus progressive multiple sclerosis), medication usage (yes/no), and, only in the case of longitudinal prediction, average baseline cognitive functioning, while sex, age and education were always entered as covariates. The threshold for including predictors was set at $P<0.10$ and predictors with $P<0.05$ were considered statistically significant.

Results

Clinical and cognitive characteristics at baseline and follow-up

The baseline characteristics for patients with multiple sclerosis and healthy controls are shown in Table 1. During the follow-up period, median EDSS increased from 3.0 at baseline to 3.5 at follow-up (P<0.01) and a total of 23/182 patients with relapsing-remitting multiple sclerosis (13%) converted to secondary progressive multiple sclerosis. Average cognitive functioning in the patient group at baseline was Z -0.88 (SD=0.89) and patients performed worst on working memory (Z = -1.29) information processing speed (Z = -1.24) and executive functioning (Z = -1.01) as published previously (Eijlers *et al.*, 2017). A total of 96/234 (41%) patients were classified as cognitively impaired at baseline. A decrease in average cognitive performance was observed in patients with multiple sclerosis between baseline and follow-up of -0.24 (SD=0.51, annualized to -0.05/yr, SD=0.11, P<0.001) as shown in Fig. 1 and Supplementary Table 1. Subdividing patients based on baseline disease phenotype showed an approximately three times faster average cognitive decline for patients with primary (-0.10/yr) and secondary progressive (-0.10/yr) multiple sclerosis than patients with relapsing-remitting multiple sclerosis (-0.03/yr, P<0.01 and P=0.03, respectively, Fig. 1B).

Classifying cognitively stable and declining patients

Classifying individual patients into "stable" or "declining" resulted in a total of 66/234 *cognitively declining* patients (28%) and 168/234 *cognitively stable* patients (72%), with only 1/66 healthy controls (2%) fulfilling this criterion for cognitive decline. This was based on the criterion that best separated cognitively declining patients from both stable patients and controls, which was a yearly rate of cognitive change < -0.25 on two or more cognitive domains (see Supplementary Table 2).The baseline demographic and clinical characteristics of the cognitively stable and declining patient groups are shown in Table 1. The cognitively declining patients showed an average cognitive change during follow-up of -0.72 (SD=0.48, annualized to -0.16/yr, SD=0.10), while the cognitively stable patients and controls had change scores of <0.01/yr by definition. The cognitively declining patient group did not differ from the cognitively stable group on baseline demographic variables and baseline cognitive functioning (average cognition Z = -0.98 vs Z = -0.84, respectively, P=0.67), but did show more severe physical disability at baseline (median EDSS 3.5 vs 3.0 respectively, P<0.01) and contained a larger proportion of progressive patients (42% vs 14% respectively, P<0.01).

Whole-brain MRI differences between cognitively stable and declining patients

Compared to the cognitively stable patients, the cognitively declining patients showed more severe structural damage at baseline on four MRI measures (see Table 1), including a higher lesion volume (13.37mL vs 8.33mL, $P<0.01$), a lower white matter integrity (0.38 vs 0.40, $P<0.01$), lower cortical grey matter volume (0.73L vs 0.76L, $P<0.01$) and deep grey matter volume (54.34mL vs 57.56mL, $P<0.01$). As a post hoc analysis, patients were split into quartiles based on the amount of whole-brain structural damage at baseline for each of these four MRI measures separately, after which the yearly rate of cognitive change was computed for each quartile. For the groups based on baseline white matter lesions and integrity, a gradual pattern was observed, with an increase in yearly rate of cognitive decline when moving from minimal to high white matter damage, which was significant after Bonferroni correction between the extreme quartiles only (Fig. 2A and 2B). For groups based on deep and cortical grey matter, a different pattern was observed with an (approximately two to three times) faster yearly rate of cognitive decline for patients in the lowest volume quartile (i.e. the most grey matter atrophy) compared to the other quartiles (Fig. 2C and 2D). However, only the differences between the lowest and second-lowest quartiles survived the stringent Bonferroni correction.

Regional MRI differences between cognitively stable and declining patients

Regional lesion volume is shown as percentages of tract volume for cognitively declining and stable patients in Fig. 3A, which was very similar between patient groups. Regional integrity damage within tracts is shown in Fig 3B as Z-scores relative to controls, showing the most severe damage in the forceps major. Atrophy of individual deep grey matter regions (Fig. 3C) and atrophy of individual cortical lobes (Fig. 3D) showed a consistent pattern of more severe regional damage in declining compared to stable patients which was significant for all regional atrophy measures, except for amygdala volume. Overall, most severe atrophy was visible in the thalamus. Comparing the effect sizes of all regional MRI measures, largest relative differences between cognitively stable and declining patients were observed for cortical volumes.

Voxel-wise MRI differences between cognitively stable and declining patients

To further localize the precise areas where brain damage differed between cognitively stable and declining patients at baseline, voxel-wise analyses were performed. The voxel-wise lesion

location comparison showed significantly more lesions in declining patients primarily in the forceps major and inferior occipitofrontal fasciculus (Fig. 4A). The voxel-wise tract integrity comparison showed most extensive involvement in the forceps major and minor, as well as in the superior and inferior longitudinal fasciculi (Fig. 4B). The deep grey matter shape analysis revealed more atrophy in cognitively declining patients near the ventricles, that is, in the medial thalamus, lateral hippocampus, the medial caudate nucleus , but also on the medial side of the left pallidum (Fig. 4C). Finally, the voxel-based morphometry analysis showed lower cortical grey matter volume in the cognitively declining patients mainly in the bilateral superior and medial temporal gyri, right inferior temporal gyrus, bilateral medial and lateral occipital lobe and bilateral inferior frontal lobe (Fig. 4D).

Prediction of cognitive impairment and decline

The cross sectional logistic regression analysis to find baseline correlates of cognitive impairment (Nagelkerke R^2 = 0.29, P<0.01) contained deep grey matter volume, multiple sclerosis phenotype (relapse onset vs primary progressive) and level of education as significant variables (Table 2). The longitudinal logistic regression analysis to predict cognitive decline using baseline *whole-brain* MRI measures (Nagelkerke R^2 = 0.22, P<0.01) included cortical grey matter volume, multiple sclerosis phenotype (relapsing-remitting vs progressive) and age as significant predictors (Table 2). The disease stage specific models showed that cognitive decline in early relapsing-remitting multiple sclerosis was predicted (Nagelkerke R^2 = 0.10) by white matter integrity only, whereas cognitive decline in late relapsing-remitting as well as in progressive multiple sclerosis was predicted by cortical atrophy only (Nagelkerke R^2 = 0.15 and 0.12, respectively). Similarly, information processing speed decline (N = 76) in early relapsing-remitting multiple sclerosis was predicted (Nagelkerke R^2 = 0.28) by deep grey matter atrophy, medication usage and baseline test score, in late relapsing-remitting multiple sclerosis (Nagelkerke R^2 = 0.39) by cortical atrophy and baseline test score and in progressive multiple sclerosis (Nagelkerke R^2 = 0.22) by cortical atrophy and symptom duration.

The four regression analyses that evaluated *regional* MRI measures are shown in Supplementary Table 3. Significant regional MRI measures in these models were occipital and temporal grey matter volume for the cortical model (Nagelkerke R^2 = 0.24, P<0.01), caudate nucleus volume for the deep grey matter model (Nagelkerke R^2 = 0.16, P<0.01), anterior thalamic radiation and superior longitudinal fasciculus integrity for the white matter

tract integrity model (Nagelkerke R^2 = 0.23, P<0.01) and forceps major and superior longitudinal fasciculus lesion percentages for the white matter tract lesion percentage model (Nagelkerke R^2 = 0.19, P<0.01). The final regional MRI regression model (Nagelkerke R^2 = 0.35, P<0.01), included anterior thalamic radiation integrity, superior longitudinal fasciculus lesion percentage, temporal cortical volume, multiple sclerosis phenotype (relapsing-remitting vs progressive) and age as significant predictors for cognitive decline (Table 2).

Discussion

In this study, we investigated cognitive decline during a five year follow-up period in a large cohort of patients with multiple sclerosis. A significant decline in cognitive functioning was observed between baseline and follow-up, which was three times faster in progressive patients compared to relapsing-remitting patients. Next, an individualized classification approach of cognitively stable and declining patients was developed, with 28% of patients classified as cognitively declining with high degree of certainty. A comparison between cognitively declining and stable patients on baseline characteristics showed that the cognitively declining patient group already demonstrated more severe structural damage at baseline, without differences on demographic or cognitive scores at that time. At baseline, the main cross-sectional MRI correlate of cognitive performance was *deep* grey matter volume, while the main baseline MRI predictor of future cognitive decline turned out to be *cortical* grey matter volume. Cognitive decline was primarily predicted by white matter integrity (global cognitive decline) and deep grey matter volume (information processing speed decline) in early multiple sclerosis, but by cortical atrophy in more advanced relapsing-remitting as well as progressive multiple sclerosis. The regression analysis that evaluated *regional* MRI measures of cognitive decline showed a substantially larger explained variance compared to the model that included whole-brain MRI predictors only, which suggests that the assessment of regional damage is valuable to pursue.

The predictive value of baseline demographic variables on cognitive decline was limited, with only age remaining as a significant predictor in the models. Although level of education strongly related to cognitive impairment in our cross sectional model, as well as in previous studies (Bonnet *et al.*, 2006; Martins Da Silva *et al.*, 2015), it did not significantly predict future cognitive decline. A high level of education is thought to reflect a high cognitive reserve and possibly protects against cognitive decline (Sumowski *et al.*, 2013). The lack of an effect of educational level on future cognitive decline could possibly suggest that this protective effect is largely depleted and most relevant in early disease stages as was also indicated in Alzheimer's disease (Stern, 2012), but could also reflect a too limited assessment of cognitive reserve, with other components such as intellectual enrichment and leisure activities not taken into account. Disease phenotype was an important clinical measure in determining future cognitive decline, with both the probability as well as average rate of cognitive decline substantially higher in progressive phenotypes compared to relapsing-remitting multiple sclerosis. Although not surprising when considering that 'progressive'

diagnoses rely on the steady progression of neurological disability (Thompson *et al.*, 2018), a faster rate of decline was not as clearly described in literature for cognitive functioning. Interestingly, baseline cognitive functioning did not differ between cognitively stable and declining patients and did not significantly predict cognitive decline, which indicates that the assessment of a patient's cognitive function might not be as valuable in predicting the likelihood of future cognitive decline as MRI measures.

More severe structural damage at baseline predicted a higher probability and rate of cognitive decline during follow-up. Especially baseline white matter integrity damage and *deep* grey matter atrophy seemed relevant in the prediction of cognitive decline in early relapsing-remitting multiple sclerosis and baseline *cortical* grey matter atrophy in late relapsing-remitting and progressive multiple sclerosis. While previous studies already showed the predictive value of global (Summers *et al.*, 2008a; Deloire *et al.*, 2011) and central atrophy (Deloire *et al.*, 2011) as well as tissue integrity (Deloire *et al.*, 2011; Filippi *et al.*, 2013) for cognitive decline, these studies included relatively early multiple sclerosis cohorts, i.e. within three years after disease onset (Summers *et al.*, 2008a), at a mean disease duration of two years (Deloire *et al.*, 2011) or four years (Filippi *et al.*, 2013). The order in which white matter integrity damage, cortical and deep grey matter atrophy occur in multiple sclerosis is a topic of great debate, with most results suggesting that white matter integrity damage and deep grey matter atrophy already start early, while cortical atrophy is more common in advanced disease stages (Kutzelnigg *et al.*, 2005; Audoin *et al.*, 2010; Bergsland *et al.*, 2012; Haider *et al.*, 2014; Schoonheim *et al.*, 2014; Steenwijk *et al.*, 2014). It could therefore be hypothesized that the accumulation of white matter integrity damage and deep grey matter atrophy are the primary drivers of cognitive decline in initial disease stages. This could explain both the predictive value of these measures for future cognitive decline in early relapsing-remitting multiple sclerosis only, as well as the strong cross sectional correlations with cognitive functioning at baseline in the entire patient group. The additional presence of cortical atrophy, which is predominant in late relapsing-remitting and progressive patients, would then potentially predispose to further cognitive decline. The almost three times faster yearly rate of cognitive decline for patients in the lowest cortical volume quartile further underscores this idea of an additive effect and could point to an acceleration of cognitive decline after patients cross a particular cortical volumetric threshold.

The underlying mechanisms responsible for cortical grey matter atrophy are still not fully understood. Several MRI studies have shown the relation between regional grey matter atrophy and damage to connected white matter tracts, possibly due to axonal transection

followed by Wallerian degeneration (Henry *et al.*, 2009; Jehna *et al.*, 2013; Steenwijk *et al.*, 2015). The central and highly connected nature of deep grey matter regions could perhaps explain why these regions are more prone to develop atrophy early in the disease, with white matter lesion accumulation already present in early multiple sclerosis (Lublin *et al.*, 2014; Azevedo *et al.*, 2018). Besides direct damage to connected tracts, other pathological processes, including local microglial activation, glutamate excitotoxicity and oxidative injury, have also been proposed to contribute to demyelination and neuroaxonal loss and ultimately grey matter atrophy (Klaver *et al.*, 2015; Mahad *et al.*, 2015; Popescu *et al.*, 2015). These mechanisms do not depend on direct damage to connected tracts, but it remains unclear whether these are the main drivers of the pronounced cortical degeneration that is thought to accelerate in progressive multiple sclerosis and whether these processes are also the main culprits in driving cognitive decline in long-term disease.

Results of the four separate regional regression analyses showed the highest explained variance for the regional cortical volume model. This indicates that the model that focused on regional cortical atrophy was better able to predict cognitive decline, compared to models that focused on either deep grey matter atrophy, white matter tract integrity or white matter tract lesion percentages. However, all four models were able to significantly predict cognitive decline, indicating that cortical volume is not the only relevant MRI predictor. The final regional model that combined different types of regional damage showed that the combination of integrity loss in the anterior thalamic radiation, lesions in the superior longitudinal fasciculus and temporal cortical atrophy was best at predicting cognitive decline. As this model showed a substantially larger explained variance compared to the model that included whole brain MRI measures, this indicates the additional value of assessing *regional* damage. Previous studies have highlighted the relation between thalamic damage and cognitive impairment in multiple sclerosis (Minagar *et al.*, 2013; Schoonheim *et al.*, 2015). However, the finding that integrity damage to the anterior thalamic radiation, which connects the cognitively relevant anterior and dorsomedial thalamic nuclei to the prefrontal cortex (Coenen *et al.*, 2012), can predict future cognitive decline was not previously shown.

A limitation of the current study is the potential (protective) effect of disease modifying treatments. Besides a reduced rate of white matter lesion accumulation, a reduced atrophy rate has now also been established for a number of treatments (Vidal-Jordana *et al.*, 2015). The protective effect on cognitive decline is less clear, although a few studies indicated such effect (Fischer *et al.*, 2000; Kappos *et al.*, 2016). The current study also indicated an effect of disease modifying treatment usage on information processing speed decline in early

relapsing-remitting patients, which should now be further investigated for the continuously expanding range of disease modifying treatment options. Another limitation is the difficulty to disentangle low premorbid grey matter volumes from disease related atrophy. Since larger premorbid brain volume was previously indicated to protect against cognitive decline, a concept called brain reserve (Sumowski *et al.*, 2013), low brain volumes could predispose to more severe cognitive decline. Although the cortical and deep grey matter volume measures used in the current study were normalized for head size, which to a large extent corrects for these effects, this effect of brain reserve could not be evaluated separately. Another possible limitation is the assumption that the curve and magnitude of a cognitive learning effect would be the same in patients and healthy controls, which still needs to be determined. A further point of improvement in future studies could be the application of a more sensitive cognitive assessment battery such as the Minimal Assessment of Cognitive Function In Multiple Sclerosis (MACFIMS) (Benedict *et al.*, 2006) or the shorter Brief International Cognitive Assessment for Multiple Sclerosis (Langdon *et al.*, 2012). While the current study primarily focused on global cognitive decline, cut-offs for clinically meaningful decline should be established for a larger number of cognitive tests, which will enable the comparison of decline predictors on individual cognitive domains. Finally, the assessment of longitudinal MRI changes is critically lacking in current literature. Combined with accurate longitudinal cognitive assessment, these studies have the potential to more precisely delineate the pathological substrates of observed changes in cognitive functioning.

To conclude, in a large cohort of patients with multiple sclerosis, we showed that over a period of five years, approximately 28% of patients demonstrated cognitive decline, which was most pronounced in progressive patients. At baseline, cognitively declining patients already showed more severe MRI measured brain damage. White matter integrity damage and deep grey matter atrophy were the main MRI predictors of future cognitive decline in early relapsing-remitting multiple sclerosis, whereas cortical atrophy was the main MRI predictor of future cognitive decline in both late relapsing-remitting as well as progressive multiple sclerosis. Future studies should now further elucidate the underlying mechanisms that lead to cortical atrophy as well as its possible role in clinical trials and patient management.

Funding

This study was supported by the Dutch MS Research Foundation, grant numbers 08-650, 13-820 and 14-358e.

References

Amato MP, Portaccio E, Goretti B, Zipoli V, Ricchiuti L, De Caro MF, *et al.* The Rao's Brief Repeatable Battery and Stroop Test: normative values with age, education and gender corrections in an Italian population. Mult Scler 2006; 12: 787-93.

Audoin B, Zaaraoui W, Reuter F, Rico A, Malikova I, Confort-Gouny S, *et al.* Atrophy mainly affects the limbic system and the deep grey matter at the first stage of multiple sclerosis. J Neurol Neurosurg Psychiatry 2010; 81: 690-5.

Azevedo CJ, Cen SY, Khadka S, Liu S, Kornak J, Shi Y, *et al.* Thalamic atrophy in multiple sclerosis: A magnetic resonance imaging marker of neurodegeneration throughout disease. Ann Neurol 2018; 83: 223-34.

Benedict RH, Cookfair D, Gavett R, Gunther M, Munschauer F, Garg N, *et al.* Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). J Int Neuropsychol Soc 2006; 12: 549-58.

Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, *et al.* Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult Scler 2017a; 23: 721-33.

Benedict RH, Zivadinov R. Risk factors for and management of cognitive dysfunction in multiple sclerosis. Nat Rev Neurol 2011; 7: 332-42.

Benedict RHB, DeLuca J, Enzinger C, Geurts JJG, Krupp LB, Rao SM. Neuropsychology of Multiple Sclerosis: Looking Back and Moving Forward. J Int Neuropsychol Soc 2017b; 23: 832-42.

Bergsland N, Horakova D, Dwyer MG, Dolezal O, Seidl ZK, Vaneckova M, *et al.* Subcortical and cortical gray matter atrophy in a large sample of patients with clinically isolated syndrome and early relapsing-remitting multiple sclerosis. AJNR Am J Neuroradiol 2012; 33: 1573-8.

Bonnet MC, Deloire MS, Salort E, Dousset V, Petry KG, Brochet B, *et al.* Evidence of cognitive compensation associated with educational level in early relapsing-remitting multiple sclerosis. J Neurol Sci 2006; 251: 23-8.

Boringa JB, Lazeron RH, Reuling IE, Ader HJ, Pfennings L, Lindeboom J, *et al.* The brief repeatable battery of neuropsychological tests: normative values allow application in multiple sclerosis clinical practice. Mult Scler 2001; 7: 263-7.

Brand N, Jolles J. Information processing in depression and anxiety. Psychol Med 1987; 17: 145-53.

Buschke H. Selective Reminding for Analysis of Memory and Learning. J Verb Learn Verb Be 1973; 12: 543-50.

Camp SJ, Stevenson VL, Thompson AJ, Ingle GT, Miller DH, Borras C, *et al.* A longitudinal study of cognition in primary progressive multiple sclerosis. Brain 2005; 128: 2891-8.

Chard DT, Jackson JS, Miller DH, Wheeler-Kingshott CA. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 2010; 32: 223-8.

Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. Lancet Neurol 2008; 7: 1139-51.

Coenen VA, Panksepp J, Hurwitz TA, Urbach H, Madler B. Human medial forebrain bundle (MFB) and anterior thalamic radiation (ATR): imaging of two major subcortical pathways and the dynamic balance of opposite affects in understanding depression. J Neuropsychiatry Clin Neurosci 2012; 24: 223-36.

Dalton CM, Bodini B, Samson RS, Battaglini M, Fisniku LK, Thompson AJ, *et al.* Brain lesion location and clinical status 20 years after a diagnosis of clinically isolated syndrome suggestive of multiple sclerosis. Mult Scler 2012; 18: 322-8.

Damjanovic D, Valsasina P, Rocca MA, Stromillo ML, Gallo A, Enzinger C, *et al.* Hippocampal and Deep Gray Matter Nuclei Atrophy Is Relevant for Explaining Cognitive Impairment in MS: A Multicenter Study. AJNR Am J Neuroradiol 2017; 38: 18-24.
Deloire MS, Ruet A, Hamel D, Bonnet M, Dousset V, Brochet B. MRI predictors of cognitive outcome in early multiple sclerosis. Neurology 2011; 76: 1161-7.
Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. Nat Rev Immunol 2015; 15: 545-58.
Douaud G, Smith S, Jenkinson M, Behrens T, Johansen-Berg H, Vickers J, *et al.* Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. Brain 2007; 130: 2375-86.
Dutta R, Trapp BD. Mechanisms of neuronal dysfunction and degeneration in multiple sclerosis. Prog Neurobiol 2011; 93: 1-12.
Eijlers AJ, Meijer KA, Wassenaar TM, Steenwijk MD, Uitdehaag BM, Barkhof F, *et al.* Increased default-mode network centrality in cognitively impaired multiple sclerosis patients. Neurology 2017; 88: 952-60.
Filippi M, Preziosa P, Copetti M, Riccitelli G, Horsfield MA, Martinelli V, *et al.* Gray matter damage predicts the accumulation of disability 13 years later in MS. Neurology 2013; 81: 1759-67.
Filli L, Hofstetter L, Kuster P, Traud S, Mueller-Lenke N, Naegelin Y, *et al.* Spatiotemporal distribution of white matter lesions in relapsing-remitting and secondary progressive multiple sclerosis. Mult Scler 2012; 18: 1577-84.
Fischer JS, Priore RL, Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, *et al.* Neuropsychological effects of interferon beta-1a in relapsing multiple sclerosis. Multiple Sclerosis Collaborative Research Group. Ann Neurol 2000; 48: 885-92.
Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage 2001; 14: 21-36.
Haider L, Simeonidou C, Steinberger G, Hametner S, Grigoriadis N, Deretzi G, *et al.* Multiple sclerosis deep grey matter: the relation between demyelination, neurodegeneration, inflammation and iron. J Neurol Neurosurg Psychiatry 2014; 85: 1386-95.
Henry RG, Shieh M, Amirbekian B, Chung S, Okuda DT, Pelletier D. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. J Neurol Sci 2009; 282: 61-6.
Iverson GL. Interpreting change on the WAIS-III/WMS-III in clinical samples. Arch Clin Neuropsychol 2001; 16: 183-91.
Jehna M, Langkammer C, Khalil M, Fuchs S, Reishofer G, Fazekas F, *et al.* An exploratory study on the spatial relationship between regional cortical volume changes and white matter integrity in multiple sclerosis. Brain Connect 2013; 3: 255-64.
Kappos L, Edan G, Freedman MS, Montalban X, Hartung HP, Hemmer B, *et al.* The 11-year long-term follow-up study from the randomized BENEFIT CIS trial. Neurology 2016; 87: 978-87.
Klaver R, Popescu V, Voorn P, Galis-de Graaf Y, van der Valk P, de Vries HE, *et al.* Neuronal and axonal loss in normal-appearing gray matter and subpial lesions in multiple sclerosis. J Neuropathol Exp Neurol 2015; 74: 453-8.
Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983; 33: 1444-52.
Kutzelnigg A, Lucchinetti CF, Stadelmann C, Bruck W, Rauschka H, Bergmann M, *et al.* Cortical demyelination and diffuse white matter injury in multiple sclerosis. Brain 2005; 128: 2705-12.

Langdon DW, Amato MP, Boringa J, Brochet B, Foley F, Fredrikson S, *et al.* Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). Mult Scler 2012; 18: 891-8.
Louapre C, Perlbarg V, Garcia-Lorenzo D, Urbanski M, Benali H, Assouad R, *et al.* Brain networks disconnection in early multiple sclerosis cognitive deficits: an anatomofunctional study. Hum Brain Mapp 2014; 35: 4706-17.
Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson AJ, *et al.* Defining the clinical course of multiple sclerosis The 2013 revisions. Neurology 2014; 83: 278-86.
Mahad DH, Trapp BD, Lassmann H. Pathological mechanisms in progressive multiple sclerosis. Lancet Neurol 2015; 14: 183-93.
Martins Da Silva A, Cavaco S, Moreira I, Bettencourt A, Santos E, Pinto C, *et al.* Cognitive reserve in multiple sclerosis: Protective effects of education. Mult Scler 2015; 21: 1312-21.
Meijer KA, Eijlers AJC, Douw L, Uitdehaag BMJ, Barkhof F, Geurts JJG, *et al.* Increased connectivity of hub networks and cognitive impairment in multiple sclerosis. Neurology 2017; 88: 2107-14.
Minagar A, Barnett MH, Benedict RH, Pelletier D, Pirko I, Sahraian MA, *et al.* The thalamus and multiple sclerosis: modern views on pathologic, imaging, and clinical aspects. Neurology 2013; 80: 210-9.
Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 2011; 56: 907-22.
Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, *et al.* Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011; 69: 292-302.
Popescu V, Battaglini M, Hoogstrate WS, Verfaillie SC, Sluimer IC, van Schijndel RA, *et al.* Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. Neuroimage 2012; 61: 1484-94.
Popescu V, Klaver R, Voorn P, Galis-de Graaf Y, Knol DL, Twisk JW, *et al.* What drives MRI-measured cortical atrophy in multiple sclerosis? Mult Scler 2015; 21: 1280-90.
Rao SM. A Manual for the Brief Repeatable Battery of Neuropsychological Tests in Multiple Sclerosis: Medical College of Wisconsin. Milwaukee: Medical College of Wisconsin. 1990.
Rocca MA, Riccitelli GC, Meani A, Pagani E, Del Sette P, Martinelli V, *et al.* Cognitive reserve, cognition, and regional brain damage in MS: A 2 -year longitudinal study. Mult Scler 2018: 1352458517750767.
Schoonheim MM, Hulst HE, Brandt RB, Strik M, Wink AM, Uitdehaag BM, *et al.* Thalamus structure and function determine severity of cognitive impairment in multiple sclerosis. Neurology 2015; 84: 776-83.
Schoonheim MM, Vigeveno RM, Rueda Lopes FC, Pouwels PJ, Polman CH, Barkhof F, *et al.* Sex-specific extent and severity of white matter damage in multiple sclerosis: implications for cognitive decline. Hum Brain Mapp 2014; 35: 2348-58.
Smith A. The Symbol Digits Modalities Test Manual, revised. Los Angeles: Western Psychological Services; 1982.
Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, *et al.* Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 2006; 31: 1487-505.
Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, *et al.* Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 2002; 17: 479-89.
Steenwijk MD, Daams M, Pouwels PJ, Balk LJ, Tewarie PK, Killestein J, *et al.* What explains gray matter atrophy in long-standing multiple sclerosis? Radiology 2014; 272: 832-42.

Steenwijk MD, Daams M, Pouwels PJ, L JB, Tewarie PK, Geurts JJ, *et al.* Unraveling the relationship between regional gray matter atrophy and pathology in connected white matter tracts in long-standing multiple sclerosis. Hum Brain Mapp 2015; 36: 1796-807.
Steenwijk MD, Geurts JJ, Daams M, Tijms BM, Wink AM, Balk LJ, *et al.* Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain 2016; 139: 115-26.
Steenwijk MD, Pouwels PJ, Daams M, van Dalen JW, Caan MW, Richard E, *et al.* Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). Neuroimage Clin 2013; 3: 462-9.
Stern Y. Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol 2012; 11: 1006-12.
Stroop JR. Studies of Interference in Serial Verbal Reactions (Reprinted from Journal Experimental-Psychology, Vol 18, Pg 643-662, 1935). J Exp Psychol Gen 1992; 121: 15-23.
Summers M, Fisniku L, Anderson V, Miller D, Cipolotti L, Ron M. Cognitive impairment in relapsing-remitting multiple sclerosis can be predicted by imaging performed several years earlier. Mult Scler 2008a; 14: 197-204.
Summers M, Swanton J, Fernando K, Dalton C, Miller DH, Cipolotti L, *et al.* Cognitive impairment in multiple sclerosis can be predicted by imaging early in the disease. J Neurol Neurosurg Psychiatry 2008b; 79: 955-8.
Sumowski JF, Benedict R, Enzinger C, Filippi M, Geurts JJ, Hamalainen P, *et al.* Cognition in multiple sclerosis: State of the field and priorities for the future. Neurology 2018; 90: 278-88.
Sumowski JF, Rocca MA, Leavitt VM, Riccitelli G, Comi G, DeLuca J, *et al.* Brain reserve and cognitive reserve in multiple sclerosis: what you've got and how you use it. Neurology 2013; 80: 2186-93.
Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018; 17: 162-73.
Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. The Concept Shifting Test: adult normative data. Psychol Assess 2006; 18: 424-32.
Vidal-Jordana A, Sastre-Garriga J, Rovira A, Montalban X. Treating relapsing-remitting multiple sclerosis: therapy effects on brain atrophy. J Neurol 2015; 262: 2617-26.
Wattjes MP, Rovira A, Miller D, Yousry TA, Sormani MP, de Stefano MP, *et al.* Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis--establishing disease prognosis and monitoring patients. Nat Rev Neurol 2015; 11: 597-606.
Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. Neuroimage 2014; 92: 381-97.
Zivadinov R, Sepcic J, Nasuelli D, De Masi R, Bragadin LM, Tommasi MA, *et al.* A longitudinal study of brain atrophy and cognitive disturbances in the early phase of relapsing-remitting multiple sclerosis. J Neurol Neurosurg Psychiatry 2001; 70: 773-80.

Figure legends

Figure 1. Cognitive change over time between baseline and follow-up.

A: To evaluate changes in cognitive functioning, a practice adjusted reliable change index score was computed with the learning effect based on changes observed in the healthy control group. B: To obtain the *yearly* cognitive change scores, the individual cognitive domain reliable change index scores were divided by the individual subject's interval duration and then averaged across domains. Error bars reflect standard error of the mean. Abbreviations: MS = multiple sclerosis, HC = healthy controls, RCI = reliable change index, RRMS = relapsing-remitting multiple sclerosis, SPMS = secondary-progressive multiple sclerosis, PPMS = primary-progressive multiple sclerosis. * Significantly different between groups.

Figure 2. Relation between baseline structural damage and yearly cognitive decline.

To further delineate the relation between baseline structural damage and the yearly rate of cognitive change during follow-up, the patient group was split into four quartiles for each MRI measure separately based on the amount of damage at baseline. Four each quartile, the average yearly rate of cognitive change was computed. A: Relation between baseline lesion volume and yearly rate of cognitive change. B: Relation between baseline white matter integrity and yearly rate of cognitive change. C: Relation between baseline deep grey matter volume and yearly rate of cognitive change. D: Relation between baseline cortical grey matter volume and yearly rate of cognitive change. Abbreviations: L = liter, mL = milliliter, FA = fractional anisotropy, RCI = reliable change index. Note: value ranges for each quartile represent rounded numbers, not the exact thresholds used to demarcate groups. * Significantly different between groups.

Figure 3. Baseline regional differences between cognitively stable and declining patients.

Regional MRI damage effect sizes at baseline were compared between cognitively stable and declining patients. A: Percentage of individual white matter tracts that was occupied by lesions in cognitively declining and stable patients. B: Integrity damage of individual white matter tracts in cognitively declining and stable patients. C: Atrophy in individual deep grey matter regions in cognitively declining and stable patients. D: Lobar cortical grey matter atrophy in cognitively declining and stable patients. Abbreviations: HC = healthy controls, Ant. = anterior, Inf. = inferior, Sup. = superior, Rad. = radiation, Hip = hippocampus. * Significantly different between cognitively stable and declining patients.

Figure 4. Baseline voxel-wise differences between cognitively stable and declining patients.
Voxel-wise comparisons were performed to more precisely localize baseline differences between cognitively stable and declining patients. A: Differences in lesion location between cognitively stable and declining patients. B: Differences in white matter tract integrity between cognitively stable and declining patients. C: Differences in deep grey matter shape between cognitively stable and declining patients. D: Differences in regional cortical density between cognitively stable and declining patients. Abbreviations: L-Tha = left thalamus, R-Tha = right thalamus, L-Hip = left hippocampus, R-Hip = right hippocampus, L-Amy = left amygdala, R-Amy = right amygdala, L-Cau = left caudate, R-Cau = right caudate, L-Put = left putamen, R-Put = right putamen, L-Pal = left pallidum, R-Pal = right pallidum.

Tables

Table 1. Baseline demographics, clinical and MRI characteristics

	Healthy controls (n=60)	Multiple sclerosis (n=234)	Cognitively stable (n=168)	Cognitively declining (n=66)
Demographics				
Male, n (%)	29 (48%)	**75 (32%)**$^{+}$	52 (31%)	23 (35%)
Age (yr)	46.45 (9.91)	47.61 (11.02)	46.77 (11.02)	49.77 (10.80)
Level of education[a]	6 (1-7)	**5 (1-7)**$^{+}$	5 (1-7)	4 (1-7)
Symptom duration (yr)		14.77 (8.43)	14.29 (8.39)	15.99 (8.47)
EDSS[a]		3 (0-8)	3 (0-8)	**3.5 (0-8)***
Disease phenotype (RRMS/SPMS/PPMS)		182/33/19	144/15/9	**38/18/10***
White matter damage				
Lesion volume (mL)[c]		9.22 (4.89-18.75)	8.33 (4.48-15.60)	**13.37 (7.20-22.82)***
White matter integrity	0.42 (0.02)	**0.40 (0.03)**$^{+}$	0.40 (0.02)	**0.38 (0.03)***
Brain volumes[b]				
Brain volume (L)	1.51 (0.06)	**1.46 (0.08)**$^{+}$	1.47 (0.07)	**1.42 (0.09)***
White matter volume (L)	0.70 (0.03)	**0.67 (0.04)**$^{+}$	0.67 (0.03)	0.66 (0.04)
Cortical grey matter volume (L)	0.78 (0.05)	**0.75 (0.05)**$^{+}$	0.76 (0.05)	**0.73 (0.05)***
Deep grey matter volume (mL)	62.70 (3.51)	**56.65 (6.34)**$^{+}$	57.56 (5.89)	**54.34 (6.89)***
Cognition (Z-scores vs healthy controls)				
Executive functioning	0.00 (0.76)	-1.01 (1.67)$^{+}$	-0.95 (1.73)	-1.17 (1.48)
Verbal memory	0.00 (0.91)	-0.63 (1.15)$^{+}$	-0.63 (1.15)	-0.62 (1.16)
Inf. Processing speed	0.00 (1.00)	-1.24 (1.29)$^{+}$	-1.13 (1.26)	-1.52 (1.33)
Verbal fluency	0.00 (1.00)	-0.55 (0.96)$^{+}$	-0.56 (0.94)	-0.52 (1.01)
Visuospatial memory	0.00 (0.94)	-0.76 (1.22)$^{+}$	-0.79 (1.26)	-0.70 (1.13)
Working memory	0.00 (0.85)	-1.29 (1.49)$^{+}$	-1.24 (1.47)	-1.42 (1.53)
Attention	0.00 (0.65)	-0.71 (1.05)$^{+}$	-0.67 (0.99)	-0.82 (1.21)
Average cognition	0.00 (0.48)	-0.88 (0.89)$^{+}$	-0.84 (0.89)	-0.98 (0.89)

Baseline demographic, clinical and MRI characteristics were compared between patients and controls (left) and between cognitively stable and declining patients (right). All values represent means and standard deviations, unless otherwise denoted. Abbreviations: yr = year, EDSS = Expanded Disability Status Scale, RRMS = relapsing-remitting multiple sclerosis, SPMS = secondary-progressive multiple sclerosis, PPMS = primary-progressive multiple sclerosis. [a] Median and range, [b] Reported brain volumes are normalized for head size, [c] median and interquartile range. $^{+}$ Significant difference between patients with multiple sclerosis and healthy controls at $P<0.05$. * Significant difference between cognitively stable and declining patients at $P<0.05$.

Tables

Table 2. Logistic regression analysis for prediction of cognitive impairment and decline.

	Model			Predictor		
	Nagelkerke R^2	**Chi-square**	***P***	**B (S.E.)**	**Wald**	***P***
Predicting cognitive impairment (whole-brain MRI measures)	**0.29**	**55.71**	**<0.01**			
Deep grey matter volume (mL)				-0.15 (0.03)	26.37	**<0.01**
Level of education (1-7)				-0.29 (0.09)	9.63	**<0.01**
MS phenotype (1: Relapse onset, 2: Primary progressive)				1.48 (0.65)	5.22	**0.02**
Sex (1: Female, 2: Male)				0.37 (0.33)	1.29	0.26
Age (yr)				-0.01 (0.02)	0.21	0.65
Predicting future cognitive decline (whole-brain MRI measures)	**0.22**	**37.92**	**<0.01**			
Cortical grey matter volume (L)				-16.91 (4.38)	14.88	**<0.01**
MS phenotype (1: Relapsing-remitting, 2: Progressive)				1.46 (0.43)	11.30	**<0.01**
Age (yr)				-0.05 (0.02)	5.01	**0.03**
Average cognition (Z-score)				0.37 (0.22)	2.79	0.10
Level of education (1-7)				-0.10 (0.10)	1.04	0.31
Sex (1: Female, 2: Male)				-0.07 (0.35)	0.04	0.84
Predicting future cognitive decline (regional MRI measures)	**0.35**	**64.58**	**<0.01**			
MS phenotype (1: Relapsing-remitting, 2: Progressive)				1.82 (0.50)	13.55	**<0.01**
Anterior thalamic radiation integrity (FA)				-50.36 (16.05)	9.85	**<0.01**
Superior longitudinal fasciculus lesions (%)				-0.60 (0.20)	8.98	**<0.01**
Age (yr)				-0.06 (0.02)	6.53	**0.01**
Temporal cortical volume (mL)				-0.05 (0.02)	5.56	**0.02**
Average cognition (Z-score)				0.46 (0.25)	3.40	0.07
Occipital cortical volume (mL)				-0.05 (0.03)	3.27	0.07
Forceps major lesions (%)				0.13 (0.07)	3.15	0.08
Superior longitudinal fasciculus integrity (FA)				26.30 (15.33)	2.94	0.09
Level of education (1-7)				-0.10 (0.11)	0.85	0.36
Sex (1: Female, 2: Male)				0.22 (0.43)	0.25	0.62

Logistic regressions analyses with backward selection were performed to predict cognitive impairment at baseline for comparative purposes and to predict cognitive decline during follow-up using baseline measures. Note: average cognition was only included in the models to predict (future) cognitive decline to avoid circularity. The threshold for including predictors was set at $P<0.10$ and predictors with $P<0.05$ were considered statistically significant. Abbreviations: MS = multiple sclerosis, B = predictor specific b-value, S.E. = standard error, yr = year, mL = milliliter, L = liter, FA = fractional anisotropy.

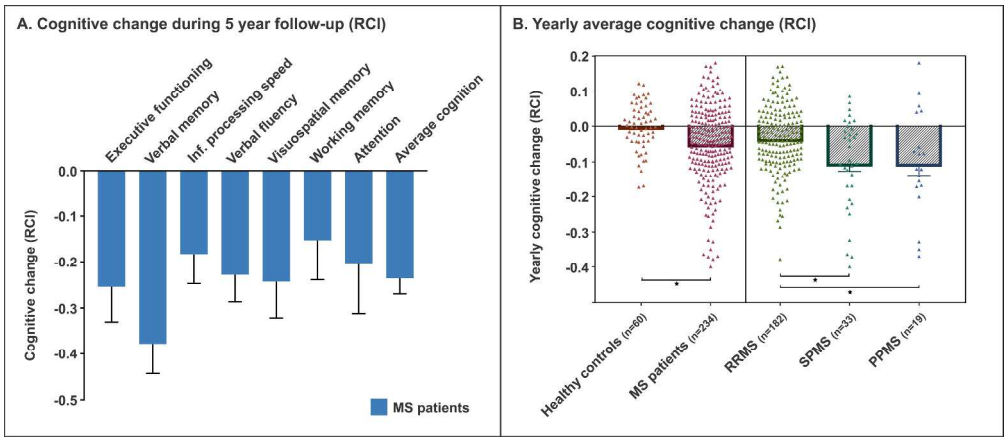

Figure 1. Cognitive change over time between baseline and follow-up.‼ ┼ A: To evaluate changes in cognitive functioning, a practice adjusted reliable change index score was computed with the learning effect based on changes observed in the healthy control group. B: To obtain the yearly cognitive change scores, the individual cognitive domain reliable change index scores were divided by the individual subject's interval duration and then averaged across domains. Error bars reflect standard error of the mean. Abbreviations: MS = multiple sclerosis, HC = healthy controls, RCI = reliable change index, RRMS = relapsing-remitting multiple sclerosis, SPMS = secondary-progressive multiple sclerosis, PPMS = primary-progressive multiple sclerosis. * Significantly different between groups.

360x155mm (300 x 300 DPI)

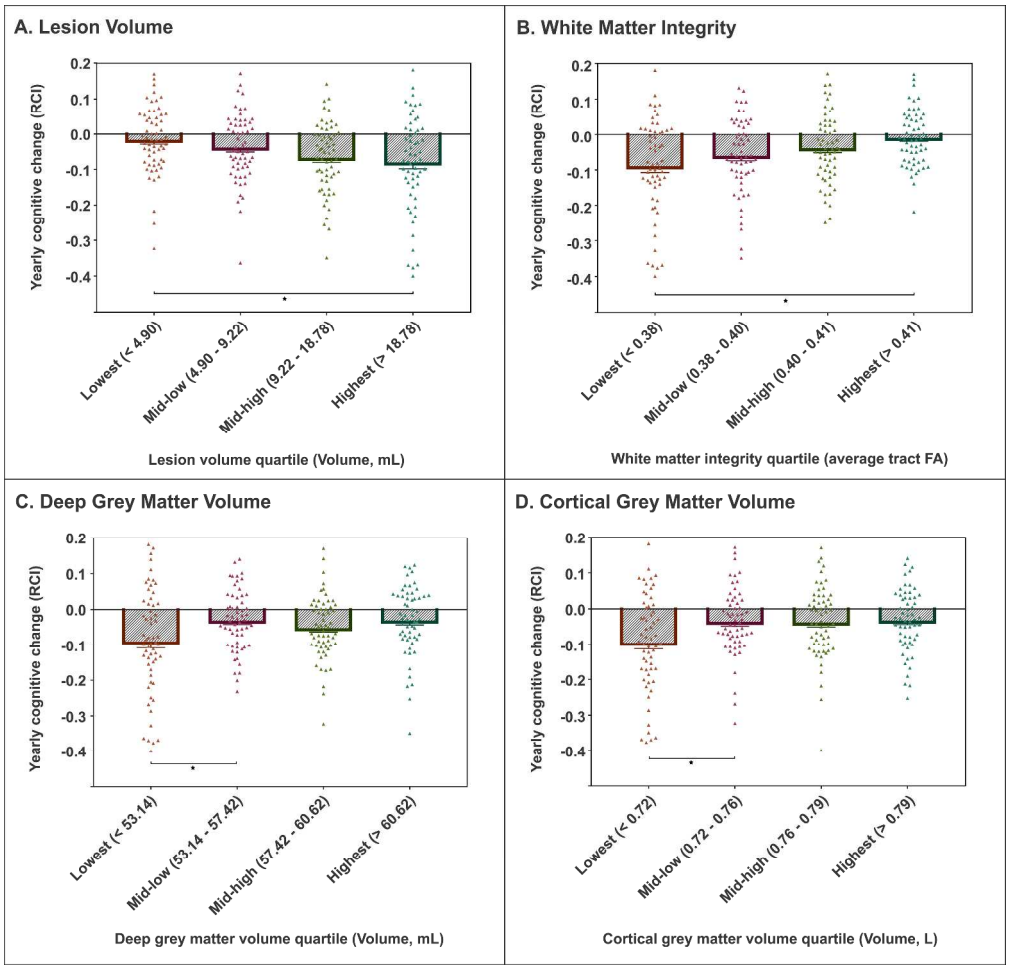

Figure 2. Relation between baseline structural damage and yearly cognitive decline.
To further delineate the relation between baseline structural damage and the yearly rate of cognitive change during follow-up, the patient group was split into four quartiles for each MRI measure separately based on the amount of damage at baseline. Four each quartile, the average yearly rate of cognitive change was computed. A: Relation between baseline lesion volume and yearly rate of cognitive change. B: Relation between baseline white matter integrity and yearly rate of cognitive change. C: Relation between baseline deep grey matter volume and yearly rate of cognitive change. D: Relation between baseline cortical grey matter volume and yearly rate of cognitive change. Abbreviations: L = liter, mL = milliliter, FA = fractional anisotropy, RCI = reliable change index. Note: value ranges for each quartile represent rounded numbers, not the exact thresholds used to demarcate groups. * Significantly different between groups.

479x459mm (300 x 300 DPI)

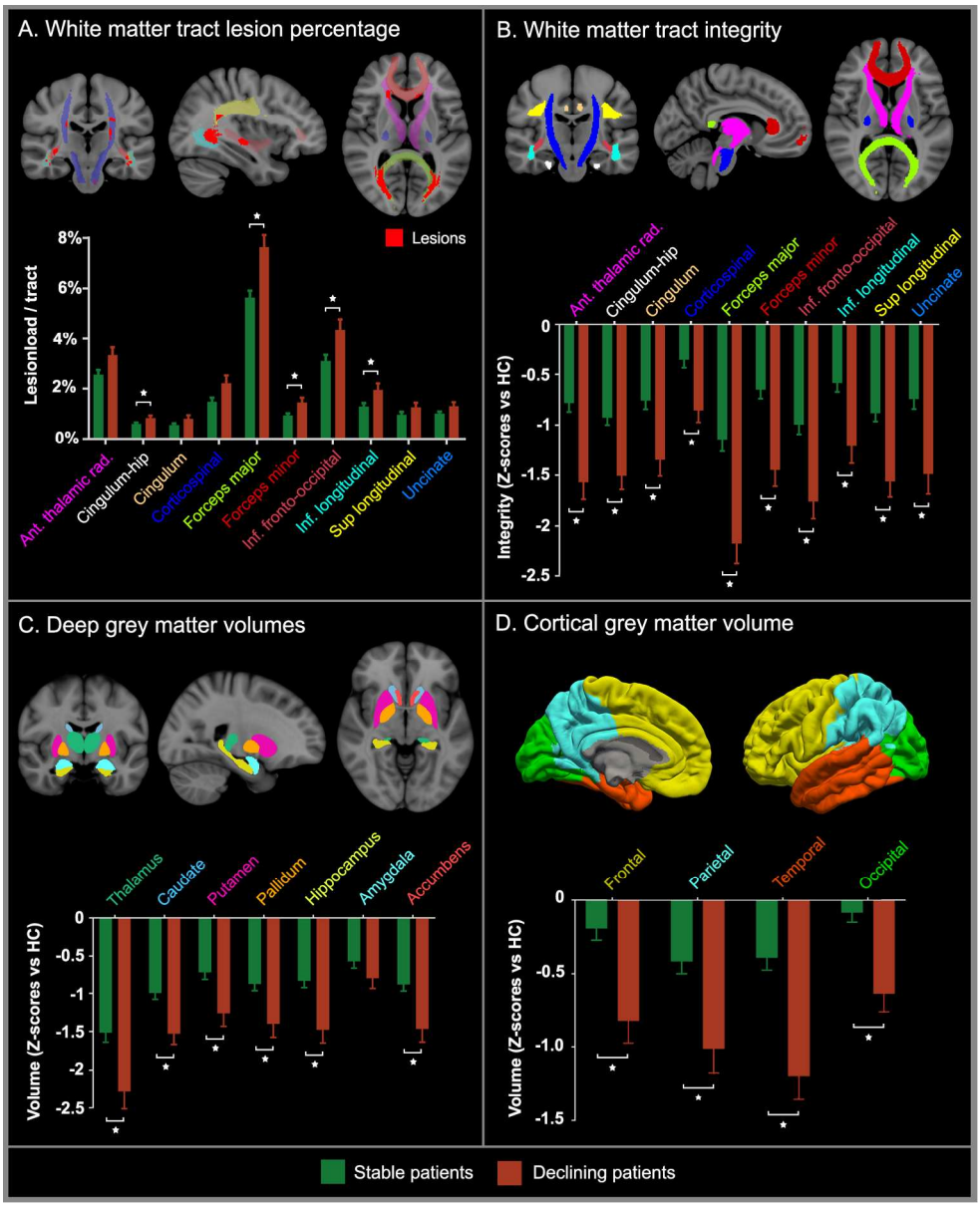

Figure 3. Baseline regional differences between cognitively stable and declining patients.
Regional MRI damage effect sizes at baseline were compared between cognitively stable and declining patients. A: Percentage of individual white matter tracts that was occupied by lesions in cognitively declining and stable patients. B: Integrity damage of individual white matter tracts in cognitively declining and stable patients. C: Atrophy in individual deep grey matter regions in cognitively declining and stable patients. D: Lobar cortical grey matter atrophy in cognitively declining and stable patients. Abbreviations: HC = healthy controls, Ant. = anterior, Inf. = inferior, Sup. = superior, Rad. = radiation, Hip = hippocampus. * Significantly different between cognitively stable and declining patients.

171x209mm (300 x 300 DPI)

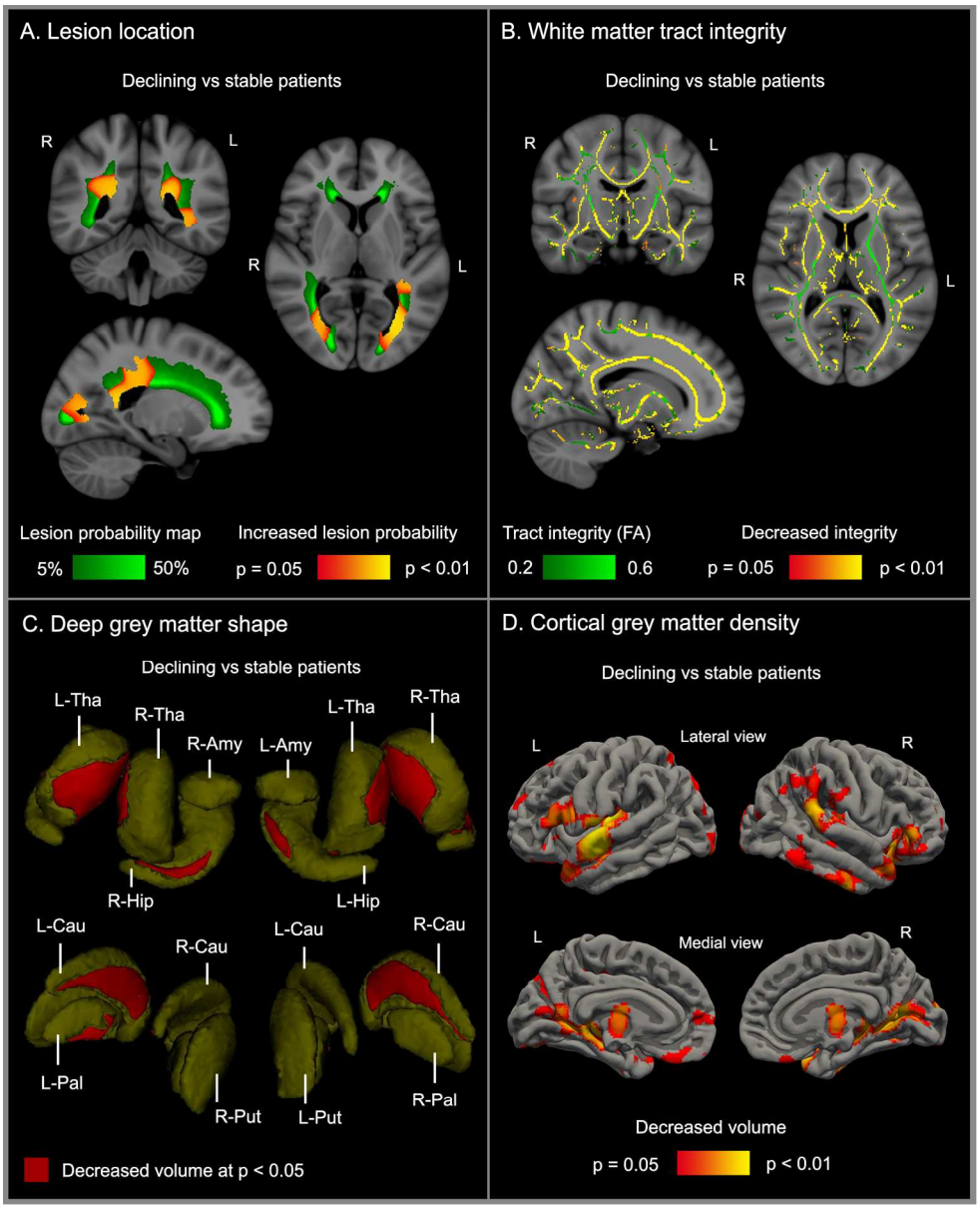

Figure 4. Baseline voxel-wise differences between cognitively stable and declining patients. Voxel-wise comparisons were performed to more precisely localize baseline differences between cognitively stable and declining patients. A: Differences in lesion location between cognitively stable and declining patients. B: Differences in white matter tract integrity between cognitively stable and declining patients. C: Differences in deep grey matter shape between cognitively stable and declining patients. D: Differences in regional cortical density between cognitively stable and declining patients. Abbreviations: L-Tha = left thalamus, R-Tha = right thalamus, L-Hip = left hippocampus, R-Hip = right hippocampus, L-Amy = left amygdala, R-Amy = right amygdala, L-Cau = left caudate, R-Cau = right caudate, L-Put = left putamen, R-Put = right putamen, L-Pal = left pallidum, R-Pal = right pallidum.

170x209mm (300 x 300 DPI)

Supplementary Tables

Supplementary Table 1. Cognitive functioning at follow-up and reliable change index

	HC (n=60)	MS (n=234)	MS-Stable (n=168)	MS-Declining (n=66)
Follow-up cognition (Z-scores)				
Executive functioning	0.13 (0.84)	-1.11 (1.69)$^{+}$	-0.86 (1.58)	-1.75 (1.80)*
Verbal memory	0.27 (0.89)	-0.70 (1.21)$^{+}$	-0.55 (1.18)	-1.09 (1.20)
Inf. Processing speed	0.27 (1.05)	-1.13 (1.26)$^{+}$	-0.91 (1.22)	-1.68 (1.22)*
Verbal fluency	0.15 (1.15)	-0.61 (1.06)$^{+}$	-0.52 (1.03)	-0.85 (1.12)
Visuospatial memory	-0.07 (0.90)	-1.10 (1.24)$^{+}$	-0.94 (1.20)	-1.50 (1.25)
Working memory	-0.06 (0.77)	-1.47 (1.63)$^{+}$	-1.17 (1.38)	-2.23 (1.95)*
Attention	0.09 (0.67)	-0.74 (1.28)$^{+}$	-0.55 (1.07)	-1.21 (1.62)*
Average cognition	0.11 (0.53)	-0.98 (0.95)$^{+}$	-0.77 (0.87)	-1.50 (0.96)*
Yearly rate of cognitive change (RCI)				
Executive functioning	0.00 (0.11)	-0.06 (0.25)	-0.01 (0.22)	-0.16 (0.30)*
Verbal memory	0.00 (0.16)	-0.09 (0.22)$^{+}$	-0.05 (0.18)	-0.19 (0.27)*
Inf. Processing speed	0.00 (0.19)	-0.04 (0.21)	-0.01 (0.19)	-0.10 (0.22)
Verbal fluency	0.00 (0.17)	-0.05 (0.19)	-0.02 (0.17)	-0.11 (0.24)
Visuospatial memory	0.01 (0.17)	-0.05 (0.26)	-0.01 (0.22)	-0.16 (0.32)*
Working memory	0.01 (0.17)	-0.03 (0.28)	0.04 (0.24)	-0.20 (0.31)*
Attention	0.00 (0.11)	-0.04 (0.35)	0.01 (0.27)	-0.18 (0.47)*
Average cognition	0.00 (0.07)	-0.05 (0.11)$^{+}$	-0.01 (0.08)	-0.16 (0.10)*

Cognitive functioning at follow-up are expressed as Z-scores compared to the healthy control group at baseline as the reference group. All values represent means and standard deviations. Abbreviations: HC: healthy controls, MS: multiple sclerosis, RCI: reliable change index. $^{+}$ Signicantly different between patients with multiple sclerosis and HC. * Signicantly different between cognitively declining and stable patients.

Supplementary Tables

Supplementary Table 2. Different cut-offs for classifying cognitively declining patients

	RCI < -0.20/yr on 2 domains		RCI < -0.25/yr on 2 domains		Average RCI < -0.10/yr		Average RCI < -0.15/yr	
Patients classified as cognitively declining	95 (41%)		66 (28%)		66 (28%)		34 (15%)	
Controls classified as cognitively declining	7 (12%)		1 (2%)		4 (7%)		2 (3%)	
	Cognitively stable (n=139)	Cognitively declining (n=95)	Cognitively stable (n=168)	Cognitively declining (n=66)	Cognitively stable (n=168)	Cognitively declining (n=66)	Cognitively stable (n=200)	Cognitively declining (n=34)
Demographics								
Male (%)	31	34	31	35	32	33	32	32
Age (yr)	46.19 (10.61)	49.70 (11.32)*	46.77 (11.02)	49.77 (10.80)	46.57 (10.76)	50.28 (11.31)*	46.93 (10.72)	51.64 (12.02)*
Level of education[a]	5 (1-7)	5 (1-7)	5 (1-7)	4 (1-7)	5 (1-7)	4 (1-7)	5 (1-7)	4 (1-7)
Symptom duration (yr)	13.73 (8.16)	16.29 (8.61)*	14.29 (8.39)	15.99 (8.47)	14.03 (8.33)	16.65 (8.43)*	14.35 (8.38)	17.25 (8.40)
EDSS[a]	3 (0-8)	3.5 (0-8)*	3 (0-8)	3.5 (0-8)*	3 (0-8)	3.5 (0-8)*	3 (0-8)	3.5 (0-8)
Disease phenotype (RRMS/SPMS/PPMS)	119/12/8	63/21/11*	144/15/9	38/18/10*	142/17/9	40/16/10*	165/23/12	17/10/7*
Cognition (Z-scores vs healthy controls)								
Average cognition (baseline)	-0.80 (0.80)	-1.01 (1.00)	-0.84 (0.89)	-0.98 (0.89)	-0.89 (0.90)	-0.87 (0.88)	-0.89 (0.89)	-0.81 (0.91)
Average cognition (follow-up)	-0.69 (0.78)	-1.40 (1.02)*	-0.77 (0.87)	-1.50 (0.96)*	-0.78 (0.89)	-1.47 (0.94)*	-0.87 (0.90)	-1.62 (1.00)*
Yearly cognitive change (RCI)	0.00 (0.07)	-0.13 (0.11)*	-0.01 (0.08)	-0.16 (0.10)*	0.00 (0.07)	-0.18 (0.08)*	-0.02 (0.08)	-0.24 (0.08)*

All values represent means and standard deviations, unless otherwise denoted. Abbreviations: EDSS = Expanded Disability Status Scale, RRMS = relapsing-remitting multiple sclerosis, SPMS = secondary-progressive multiple sclerosis, PPMS = primary-progressive multiple sclerosis, yr = year, mL = milliliter, L = liter, RCI = reliable change index. [a] Median and range. * Significantly different between cognitively declining and stable patients.

Supplementary Tables

Supplementary Table 3. Logistic regression analyses using regional MRI measures.

	Model			Predictor		
	Nagelkerke R^2	Chi-square	*P*	B (S.E.)	Wald	*P*
Predicting cognitive decline (white matter tract lesions)	**0.19**	**32.59**	**<0.01**			
MS phenotype (1: Relapsing-remitting, 2: Progressive)				1.68 (0.44)	14.36	**<0.01**
Forceps major lesions (%)				0.20 (0.06)	10.78	**<0.01**
Superior longitudinal fasciculus lesions (%)				-0.28 (0.14)	4.36	**0.04**
Age (yr)				-0.02 (0.02)	1.68	0.19
Level of education (1-7)				-0.08 (0.09)	0.80	0.37
Sex (1: Female, 2: Male)				-0.26 (0.35)	0.55	0.46
Predicting cognitive decline (white matter tract integrities)	**0.23**	**40.27**	**<0.01**			
MS phenotype (1: Relapsing-remitting, 2: Progressive)				1.65 (0.46)	13.21	**<0.01**
Anterior thalamic radiation integrity (FA)				-49.93 (15.03)	11.04	**<0.01**
Superior longitudinal fasciculus integrity (FA)				30.16 (14.65)	4.24	**0.04**
Symptom duration (yr)				-0.04 (0.03)	2.68	0.10
Sex (1: Female, 2: Male)				0.47 (0.37)	1.64	0.20
Level of education (1-7)				-0.03 (0.10)	0.09	0.76
Age (yr)				-0.01 (0.02)	0.06	0.80
Predicting cognitive decline (deep grey matter volumes)	**0.16**	**26.88**	**<0.01**			
MS phenotype (1: Relapsing-remitting, 2: Progressive)				1.58 (0.42)	14.22	**<0.01**
Caudate volume (mL)				-0.38 (0.16)	5.65	**0.02**
Age (yr)				-0.02 (0.02	1.48	0.22
Level of education (1-7)				-0.06 (0.09)	0.37	0.54
Sex (1: Female, 2: Male)				-0.13 (0.34)	0.15	0.70
Predicting cognitive decline (cortical grey matter volumes)	**0.24**	**42.42**	**<0.01**			
MS phenotype (1: Relapsing-remitting, 2: Progressive)				1.58 (0.44)	12.91	**<0.01**
Temporal cortical volume (mL)				-0.05 (0.02)	7.19	**<0.01**
Occipital cortical volume (mL)				-0.06 (0.03)	5.39	**0.02**
Age (yr)				-0.04 (0.02)	4.20	**0.03**
Average cognition (Z-score)				0.44 (0.23)	3.67	0.06
Level of education (1-7)				-0.12 (0.10)	1.51	0.22
Sex (1: Female, 2: Male)				-0.02 (0.35)	0.00	0.95

Logistic regressions analyses with backward selection were performed to predict cognitive decline during follow-up using baseline regional MRI measures, with separate models for white matter tract lesion percentages, white matter tract integrities, deep grey matter volumes and cortical lobe volumes. The threshold for including predictors was set at $P<0.10$ and predictors with $P<0.05$ were considered statistically significant. Abbreviations: MS = multiple sclerosis, B = predictor specific b-value, S.E. = standard error, yr = year, mL = millilitre, L = litre, FA = fractional anisotropy.