

ORIGINAL ARTICLE

Central nervous system atrophy predicts future dynamics of disability progression in a real-world multiple sclerosis cohort

Charidimos Tsagkas^{1,2,3}  | Yvonne Naegelin¹ | Michael Amann^{1,3,4} |
Athina Papadopoulou^{1,2} | Christian Barro^{1,5}  | M. Mallar Chakravarty^{6,7,8} |
Laura Gaetano⁹ | Jens Wuerfel^{3,4} | Ludwig Kappos^{1,2} | Jens Kuhle¹ |
Cristina Granziera^{1,2,3,4} | Till Sprenger^{1,10} | Stefano Magon¹¹ | Katrin Parmar^{1,2}

¹Neurologic Clinic and Policlinic, Departments of Medicine, Clinical Research and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland

²Translational Imaging in Neurology (ThINK) Basel, Department of Medicine and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland

³Medical Image Analysis Center AG, Basel, Switzerland

⁴Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland

⁵Ann Romney Center for Neurologic Diseases, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

⁶Department of Psychiatry, McGill University, Montreal, QC, Canada

⁷Cerebral Imaging Centre, Douglas Mental Health University Institute, Verdun, QC, Canada

⁸Department of Biomedical Engineering, McGill University, Montreal, QC, Canada

⁹F. Hoffmann-La Roche Ltd, Basel, Switzerland

¹⁰Department of Neurology, DKD HELIOS Klinik Wiesbaden, Wiesbaden, Germany

¹¹Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland

Correspondence

Charidimos Tsagkas, Department of Neurology, University Hospital Basel, Petersgraben 4, CH-4031 Basel, Switzerland.

Email: charidimos.tsagkas@usb.ch

Funding information

C.T. was supported by the Swiss National Science Foundation (Grant number: 320030_156860), the *Stiftung zur Förderung der gastroenterologischen und allgemeinen klinischen Forschung*, as well as by the University of Basel (Grant numbers: 3MS1020 and 3MS1049). C.B. is supported by a postdoctoral fellowship from the Swiss National Science Foundation (P400PM_191077). C.G. is funded by the Swiss National Science Foundation (Grant number: PP00P3_176984), the *Stiftung zur Förderung der gastroenterologischen und allgemeinen klinischen Forschung*, and the Eurostar programme of Horizon 2020 (Grant number: E!113682)

Abstract

Background and purpose: In an era of individualized multiple sclerosis (MS) patient management, biomarkers for accurate prediction of future clinical outcomes are needed. We aimed to evaluate the potential of short-term magnetic resonance imaging (MRI) atrophy measures and serum neurofilament light chain (sNfL) as predictors of the dynamics of disability accumulation in relapse-onset MS.

Methods: Brain gray and white matter, thalamic, striatal, pallidal and cervical spinal cord volumes, and lesion load were measured over three available time points (mean time span 2.24 ± 0.70 years) for 183 patients (140 relapsing-remitting [RRMS] and 43 secondary-progressive MS [SPMS]; 123 female, age 46.4 ± 11.0 years; disease duration 15.7 ± 9.3 years), and their respective annual changes were calculated. Baseline sNfL was also measured at the third available time point for each patient. Subsequently, patients underwent annual clinical examinations over 5.4 ± 3.7 years including Expanded Disability Status Scale (EDSS) scoring, the nine-hole peg test and the timed 25-foot walk test.

Results: Higher annual spinal cord atrophy rates and lesion load increase predicted higher future EDSS score worsening over time in SPMS. Lower baseline thalamic volumes

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *European Journal of Neurology* published by John Wiley & Sons Ltd on behalf of European Academy of Neurology.

predicted higher walking speed worsening over time in RRMS. Lower baseline gray matter, as well as higher white matter and spinal cord atrophy rates, lesion load increase, baseline striatal volumes and baseline sNfL, predicted higher future hand dexterity worsening over time. All models showed reasonable to high prediction accuracy.

Conclusion: This study demonstrates the capability of short-term MRI metrics to accurately predict future dynamics of disability progression in a real-world relapse-onset MS cohort. The present study represents a step towards the utilization of structural MRI measurements in patient care.

KEYWORDS

atrophy, biomarkers, MRI, multiple sclerosis, prediction models

INTRODUCTION

In an era of individualized multiple sclerosis (MS) patient management, biomarkers for accurate prediction of future clinical outcomes are needed. Currently, signal intensity changes on magnetic resonance imaging (MRI) representing focal inflammatory events are the main considerations in clinical routine and guide therapeutic decisions [1]. However, numerous studies have demonstrated that metrics of diffuse central nervous system (CNS) neurodegeneration, measured on serial MRI acquisitions as volume loss over time, correlate better with clinical progression occurring within the same time frame [2,3]. Additionally, serum neurofilament light chain (sNfL) levels have been shown to be associated with disease activity as well as long-term brain and cervical spinal cord (cSC) volume loss [4].

In clinical settings, however, the value of models assessing correlations between concurrently acquired disability measures and potential disease biomarkers is limited. Individualized patient management would require prediction of the dynamics of clinical disease progression in the future. For instance, it would be meaningful to identify patients at risk of losing walking ability faster than others and to be able to predict accurately the extent of this clinical progression.

Based on the well-established strong correlation between clinical disability measures and concurrently acquired longitudinal metrics of CNS atrophy, the aim of the present study was to evaluate whether cross-sectional and short-term atrophy measures, quantified on MRI, as well as cross-sectional sNfL values, can predict future dynamics of disability accumulation in a large real-world relapse-onset MS cohort.

METHODS

Study design and participants

We analyzed selected data from a large cohort of relapsing-remitting (RR) and secondary-progressive (SP) MS patients (235 patients in total recruited at baseline) from a single center (MS Center, University Hospital, Basel) [3,5–7] in retrospective fashion. Patients were followed over a maximum of 13 years (14 annual time points). The diagnosis of MS was made in accordance with international

panel-established criteria [8]. The local ethics committee approved the study (EKBB-46/04) and all patients signed informed consent.

Procedures

All patients underwent yearly standardized neurological and neuropsychological examination, conducted by trained and certified examiners. The annual examination included Expanded Disability Status Scale (EDSS) classification (<http://www.neurostatus.org>), the nine-hole peg test with the dominant and non-dominant hand (D9HPT and ND9HPT, respectively), the timed 25-foot walk test (T25fwt) as well as the Symbol Digit Modalities Test (SDMT) and Paced Auditory Serial Addition Test (PASAT). For our purposes, clinical data only from the third available time point onwards were used for further analysis, representing the future clinical development (Table S1). Treatment was documented in each follow-up (Table S2).

In contrast, MRI data of the first three available time points for each patient were included in the image analyses representing baseline volumes/volume changes (Table S1). The use of three MRI scans for the calculation of longitudinal atrophy and lesion load change measures was selected to minimize potential variability originating from physiological (e.g. hydration status) as well as image acquisition and image segmentation factors. All MRI scans were performed using the same 1.5-T Magnetom Avanto magnetic resonance scanner (Siemens Healthineers, Erlangen, Germany). The MRI protocol included a high-resolution three-dimensional T1-weighted magnetization-prepared rapid acquisition with gradient echo sequence of the brain, acquired in sagittal orientation (TR/TI/TE = 2080/1100/3.0 ms; flip angle = 15°, 160 slices, resolution: 0.98 × 0.98 × 1 mm³), which also covered the upper cSC. Additionally, a double-spin-echo proton density/T2-weighted sequence was acquired (TR/TE1/TE2 = 3980/14/108 ms; 40 slices, 3-mm slice thickness without gap with an in-plane resolution of 1 × 1 mm²).

Patient serum samples were collected on the same day as the clinical visit and sNfL levels were measured by Simoa assay, as previously described [4]. For our purposes, the sNfL data from only the third available time point for each patient (deployed as baseline measurement similarly to previous studies [9]) were used for further analysis (Table S1).

Analysis of MRI

Lesion segmentation

All brain white matter (WM) lesions were segmented on T2-weighted and proton density images by trained expert observers according to the standard operating procedures used at the local institution for the analysis of clinical phase II and phase III trials and were then filled to correct for tissue misclassification due to MS lesions [10]. Using these segmentations, the T2-weighted lesion volume of each patient was extracted.

Volumetric quantification of CNS structures

All morphological analyses were performed on the T1-weighted brain images. Brain gray matter (GM) and WM volume were computed for each patient with the fully automated tool "SIENAX" for cross-sectional studies version 2.6 [11,12]. The volume of the deep GM nuclei, including thalamic, striatal and pallidal volumes, was estimated based on an established nomenclature [13] using the "MAGeT" algorithm [14,15] as described in previous studies [16–18]. Both quantification methods were performed on T1-weighted images after lesion-filling in order to reduce biases related to tissue misclassification and to improve image registration [19,20]. A 35-mm long cSC segment, extending roughly between the foramen magnum and the C2/C3 intervertebral disc, was analysed using "CORDIAL", as described in previous methodological and clinical studies [3,7,21].

All segmentations were visually inspected for quality and excluded from further statistical analysis in case of segmentation errors. The "SIENAX" baseline volume correction factor regarding variations in head size was used for normalizing the volumes of all CNS structures beside the cSC [12]. All analyses were performed on these normalized volumes.

Statistical analysis

The mean annual volume change rate (AVCR) of all CNS structures at the beginning of our study period was calculated as the percentage of the annualized changes using the first three available time points for every patient (Appendix S1 and Table S1). Similarly, the mean absolute annual volume change (AAVC) of T2-weighted lesion load (in mm^3) was calculated for each patient in the same scans.

To approximate a normal distribution, logarithmic transformations were performed for the EDSS, D9HPT and ND9HPT, whereas an inverse transformation and a cubic transformation were conducted for the T25fwt and PASAT, respectively.

The RRMS and SPMS patient groups were evaluated separately. We constructed linear mixed-effect regression (LMER) models predicting the patients' annually assessed clinical outcomes (EDSS, D9HPT, ND9HPT, T25fwt, SDMT and PASAT of third time point onwards over 5.4 ± 3.7 years), collected over 11 years, and entered as dependent variables using MRI metrics, with baseline NFL as an

independent variable. In the initial full model, we entered demographics (sex and age), disease duration, disease type (in models including the relapse-onset MS cohort), medication (grouped as injectable, oral or infused), and years of education (for PASAT and SDMT analysis), MRI metrics (baseline and AVCR/AAVC) and baseline sNFL values as independent variables. In the case of D9HPT and ND9HPT, contralateral thalamic, striatal and pallidal volumes were entered into the LMER model. All LMER analyses were performed using a random intercept and a random time slope for each subject to allow for within-subject and between-subject variance (Appendix S1). For each independent variable, both the main effect (corresponding to the correlation to the intercept of the respective clinical outcome) as well as the interaction term with the time variable (corresponding to the correlation to the slope of the respective clinical outcome over time) were evaluated. The interaction terms between independent variables and the time variable model how these modify the course of the outcome over time, after correcting for the main effects of the independent variables. We then used a step-down model-building approach for best model selection as proposed before [22,23], which is based on a deletion of effects from the full model using *F*-statistics and the Satterthwaite's approximation to degrees of freedom with a *p*-value threshold of 0.05. In case of significant interaction terms with time, the respective main effects were kept in the model according to the principle of marginality. The primary goal of this work was to find models that perform well in predicting the future development (or course) of important clinical outcomes and not to evaluate the utility of individual biomarkers. In addition, despite multicollinearity originating from possible moderate or high correlations between predictive variables entered in our models, multicollinearity issues do not influence precision of the models' predictions, or the goodness-of-fit statistics. Therefore, we did not account for multicollinearity issues in our analysis. After model selection, *p* values were not used for any testing of statistical hypothesis and were solely reported for completeness in Tables S1 to S5; therefore, multiple comparisons correction was not performed.

For each final LMER model, we then conducted a leave-one-out fashion cross-validation analysis. In order to measure the accuracy of our prediction models, we calculated the mean absolute error, the root-mean-square error and the mean absolute percentage error (MAPE) as well as their respective 95% confidence intervals (CIs). We then ranked the accuracy of our models according to their mean MAPE values as proposed before [24]. Specifically, a MAPE of 0%–10% was graded as high predictive accuracy, a MAPE of 10%–20% as good predictive accuracy, a MAPE of 20%–50% as reasonable predictive accuracy and a MAPE of >50% as inaccurate prediction.

All statistical analyses were conducted using R version 3.6.3 (<https://www.r-project.org/>).

RESULTS

After exclusion of all relapse-onset patients with less than three available yearly MRI sessions (52 patients in total), data from

183 MS patients (140 RRMS, 43 SPMS) were used for further statistical analysis. After three yearly MRI examinations (mean time span: 2.24 ± 0.70 years; minimum 2 years), patients received clinical follow-ups for up to 11 years (mean follow-up time 5.4 ± 3.7 years; mean number of follow-ups 6.0 ± 3.6 ; Figure 1). Baseline demographics and clinical characteristics, are described in Table 1. Baseline MRI metrics, sNfL values, AVCR and AAVC are shown in Table 2. Baseline MRI metrics and sNfL as well as AVCR are also shown in Figure 2. Trajectories of clinical scores are shown in Figure 3.

Prediction of future clinical progression

Details of all final LMER models for the EDSS, 9HPT and T25fwt are shown in detail in Tables 3 and 4, as well as in Tables 3 and 4. Interestingly, mostly baseline and short-term atrophy measures were predictive of future clinical worsening over time. Important predictors for all clinical outcomes in the RRMS, SPMS groups and the whole cohort are summarized in Table 5.

Expanded Disability Status Scale

Relapsing-remitting MS

No variables were associated with future EDSS changes over time (mean yearly increase of $\log[\text{EDSS}]$ of $0.014 \pm 3.1 \times 10^{-3}$, $p = 2.6 \times 10^{-5}$). The cross-validation analysis demonstrated good predictive accuracy with a mean MAPE of 17.0% (95% CI 14.7%–19.3%).

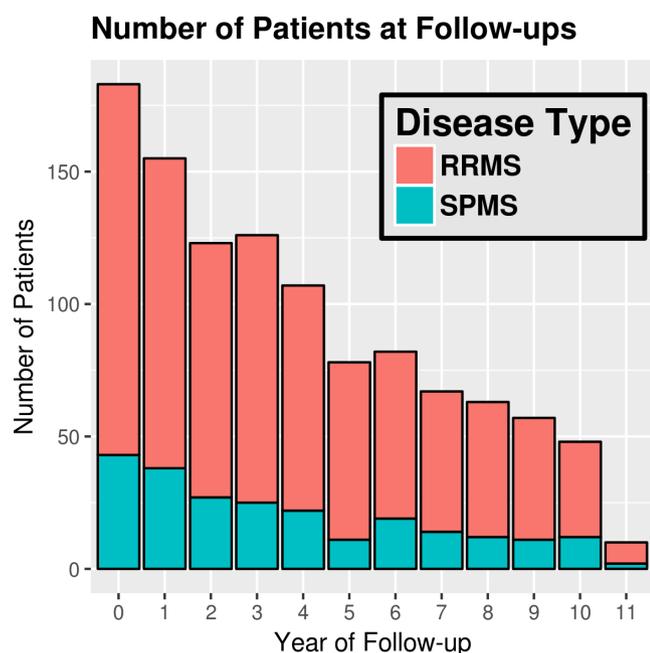


FIGURE 1 Number of patients participating in each follow-up. RRMS (red), relapsing-remitting multiple sclerosis; SPMS (turquoise), secondary progressive multiple sclerosis [Colour figure can be viewed at wileyonlinelibrary.com]

The fact that our model had good predictive accuracy using only predictors of the average patient EDSS may suggest a low variability of EDSS increase over time between patients.

Secondary progressive multiple sclerosis

Higher cSC AVCR (faster volume loss) and lesion-load AAVC (faster increase) were significantly associated with higher future EDSS worsening over time (mean yearly increase of $\log[\text{EDSS}]$ of $0.024 \pm 5.2 \times 10^{-3}$, $p = 6.7 \times 10^{-5}$). The cross-validation analysis demonstrated good predictive accuracy with a mean MAPE of 11.2% (95% CI 8.4%–14.1%).

Timed 25-foot walk test

Relapsing-remitting MS

Lower baseline thalamic volumes were significantly associated with higher future T25fwt worsening over time (mean yearly decrease of $1/\text{T25fwt}$ of $-6.4 \times 10^{-4} \pm 3.6 \times 10^{-4}$, $p = 0.095$). The cross-validation analysis demonstrated reasonable predictive accuracy, with a mean MAPE of 36.9% (95% CI 22.3%–51.6%).

Secondary progressive MS

Higher baseline GM, baseline lesion volumes, GM AVCR (faster volume loss), WM AVCR (faster volume loss) and cSC AVCR (faster volume loss) as well as lower baseline pallidal volumes were significantly associated with higher future T25fwt worsening over time (mean yearly decrease of $1/\text{T25fwt}$ of $-4.2 \times 10^{-3} \pm 1.7 \times 10^{-4}$; $p = 0.087$). The cross-validation analysis demonstrated inaccurate prediction, with a mean MAPE of 167.9% (95% CI 71.2%–264.7%).

Dominant hand nine-hole peg test

Relapsing-remitting MS

In terms of the dominant hand-dexterity, lower baseline GM volumes were significantly associated with higher future D9HPT worsening over time (mean yearly increase of $\log[\text{D9HPT}]$ of $0.012 \pm 2.0 \times 10^{-3}$, $p = 7.5 \times 10^{-7}$). The cross-validation analysis demonstrated highly accurate prediction with a mean MAPE of 5.5% (95% CI 4.8%–6.3%).

Regarding the analyses of the non-dominant hand dexterity, higher lesion load AAVC (faster increase) was significantly associated with higher future ND9HPT worsening over time (mean yearly increase of $\log[\text{ND9HPT}]$ of $0.011 \pm 1.6 \times 10^{-3}$, $p = 1.9 \times 10^{-8}$). The cross-validation analysis demonstrated highly accurate prediction with a mean MAPE of 5.7% (95% CI 4.8%–6.5%).

TABLE 1 Baseline demographics and clinical characteristics

Characteristics	Overall	RRMS	SPMS	<i>p</i>
Number of patients	183	140	43	
Baseline age, years				
Mean ± SD	46.4 ± 11.0	43.8 ± 10.2	55.0 ± 8.8	6.5 × 10 ⁻¹⁰
Range: min;max	21;70	21;70	25;69	
Women/men	123/60	99/41	24/19	0.102
Baseline disease duration, years				
Mean ± SD	15.7 ± 9.3	14.0 ± 8.7	21.3 ± 9.2	1.8 × 10 ⁻⁵
Range: min;max	2;49	2; 43	7;49	
Baseline EDSS score				
Median	3.0	2.5	5.0	5.8 × 10 ⁻¹⁵
Range: min;max	0;7.5	0;6.5	1.5;7.5	
Baseline T25fwt, s				
Mean ± SD	7.4 ± 8.3	5.6 ± 3.7	14.4 ± 14.8	3.4 × 10 ⁻¹²
Range: min;max	2.2;73.5	2.2;32.5	4.3;73.5	
Baseline D9HPT, s				
Mean ± SD	23.4 ± 13.4	20.1 ± 8.4	31.7 ± 21.2	5.6 × 10 ⁻⁶
Range: min;max	13.7;132.9	13.7;83.7	16.8;132.9	
Baseline ND9HPT, s				
Mean ± SD	25.4 ± 19.6	23.5 ± 19.2	31.6 ± 19.7	10 ⁻⁶
Range: min;max	14.5;215.5	14.5;215.5	18.7;145.0	
Baseline SDMT				
Mean ± SD	47.0 ± 13.4	48.9 ± 13.8	40.6 ± 9.9	9.6 × 10 ⁻⁵
Range: min;max	11.0;94.0	11.0;94.0	21.0;69.0	
Baseline PASAT				
Mean ± SD	45.9 ± 12.5	46.9 ± 11.9	40.7 ± 14.6	0.065
Range: min;max	2;60	2;60	7;60	
Baseline medication				
Interferon	151	116	35	0.928
Mitoxantrone	27	20	7	
Glatimer Acetate	3	2	1	
Mycophenolate mofetil	1	1	0	
Natalizumab	1	1	0	
Number of follow-ups				
Mean ± SD	6.0 ± 3.6	6.2 ± 3.6	5.5 ± 3.5	0.269
Range: min;max	1;12	1;12	1;12	

Note: Between-group comparisons for baseline demographic and clinical data were performed using Welch's two sample *t*-test and Pearson's chi-squared test with Yate's continuity correction where appropriate.

Abbreviations: D9HPT, dominant hand nine-hole peg test; EDSS, Expanded Disability Status Scale; ND9HPT, non-dominant hand nine-hole peg test; PASAT, Paced Auditory Serial Addition Test; RRMS, relapsing-remitting multiple sclerosis; SD, standard deviation; SDMT, single digit modality test; SPMS, secondary progressive multiple sclerosis; T25fwt, timed 25-foot walk test.

Secondary progressive MS

In terms of the dominant hand dexterity, no variables were associated with future D9HPT changes over time (mean yearly increase of log[D9HPT] of 0.031 ± 0.011; *p* = 0.013). The cross-validation

analysis demonstrated highly accurate prediction, with a mean MAPE of 9.2% (95% CI 7.2%–11.1%).

For the non-dominant hand dexterity, higher baseline striatal volumes, baseline NfL, WM AVCR (faster volume loss) and spinal AVCR (faster volume loss), as well as higher lesion-load AAVC (faster

TABLE 2 Baseline magnetic resonance imaging metrics and neurofilament light chain values as well as annual rates of central nervous system volume change

Characteristics	Overall	RRMS	SPMS	p
BL GMV, cm ³				
Mean ± SD	745.4 ± 61.1	756.9 ± 61.0	708.1 ± 44.7	5.6 × 10 ⁻⁸
Range: min;max	580.2;922.1	580.2;922.1	609.1;820.3	
BL WMV, cm ³				
Mean ± SD	727.8 ± 50.6	731.7 ± 50.9	715.3 ± 47.9	0.053
Range: min;max	587.0;834.7	587.0;834.7	638.3;811.3	
BL THV, cm ³				
Mean ± SD	12.9 ± 2.10	13.1 ± 2.16	12.1 ± 1.42	0.072
Range: min;max	7.17;17.2	7.17;17.2	8.54;14.7	
BL STV, cm ³				
Mean ± SD	20.8 ± 2.42	21.1 ± 2.5	19.9 ± 1.8	0.219
Range: min;max	14.7; 28.1	14.7; 28.1	16.0; 26.4	
BL PAV, cm ³				
Mean ± SD	3.16 ± 0.28	3.19 ± 0.37	3.06 ± 0.35	0.689
Range: min;max	2.17;4.46	2.17;4.19	2.48;4.46	
BL SCV, cm ³				
Mean ± SD	2.34 ± 0.32	2.39 ± 0.304	2.17 ± 0.33	3.9 × 10 ⁻⁶
Range: min;max	1.55; 3.05	1.60; 3.05	1.55; 2.87	
BL T2LV, cm ³				
Mean ± SD	6.28 ± 6.81	5.83 ± 6.17	7.73 ± 8.52	0.102
Range: min;max	0;30.7	0;24.0	0.01;30.7	
BL NfL, pg/ml				
Mean ± SD	35.9 ± 21.2	33.5 ± 19.7	43.6 ± 24.3	0.006
Range: min;max	1.3;141.7	8.5;120.1	1.3;141.7	
GMV ACR, %				
Mean ± SD	-0.25 ± 1.00	-0.27 ± 0.97	-0.17 ± 1.08	0.523
Range: min;max	-3.78;2.14	-3.78;2.13	-3.22;1.64	
WMV ACR, %				
Mean ± SD	-0.38 ± 0.98	-0.38 ± 1.01	-0.39 ± 0.87	0.952
Range: min;max	-3.69;1.92	-3.69;1.92	-2.11;1.70	
THV ACR (%)				
Mean ± SD	-0.44 ± 1.11	-0.50 ± 1.15	-0.22 ± 0.94	0.149
Range: min;max	-5.00;2.50	-5.00;2.50	-2.68;1.28	
STV ACR (%)				
Mean ± SD	-0.17 ± 0.84	-0.21 ± 0.82	-0.01 ± 0.90	0.166
Range: min;max	-2.41;2.07	-2.41;1.53	-2.39;2.07	
PAV ACR, %				
Mean ± SD	-0.81 ± 1.36	-0.77 ± 1.36	-0.94 ± 1.36	0.458
Range: min;max	-6.26;2.84	-6.26;2.84	-4.06;2.14	
SCV ACR, %				
Mean ± SD	-0.38 ± 1.25	-0.45 ± 1.19	-0.65 ± 1.45	0.355
Range: min;max	-4.45;2.62	-4.09;2.62	-4.45;2.57	
T2LV AAC, cm ³				
Mean ± SD	0.13 ± 0.42	0.14 ± 0.41	0.11 ± 0.45	0.449
Range: min;max	-0.88;1.67	-0.88;1.67	-0.50;1.43	

Note: Between-group comparisons for baseline MRI metrics and NfL values were performed using analysis of covariance after correcting for sex, age and disease duration. Between-group comparisons for ACR of MRI metrics were performed using analysis of covariance after correcting for sex, age, disease duration and baseline values.

Abbreviations: AAC, annual absolute change; ACR, annual change rate; BL, baseline; GMV, cerebral gray matter volume; MRI, magnetic resonance imaging; NfL, neurofilament light chain; PAV, pallidal volume; RRMS, relapsing-remitting multiple sclerosis; SCV, spinal cord volume; SD, standard deviation; SPMS, secondary progressive multiple sclerosis; STV, striatal volume; THV, thalamic volume; WMV, cerebral white matter volume.

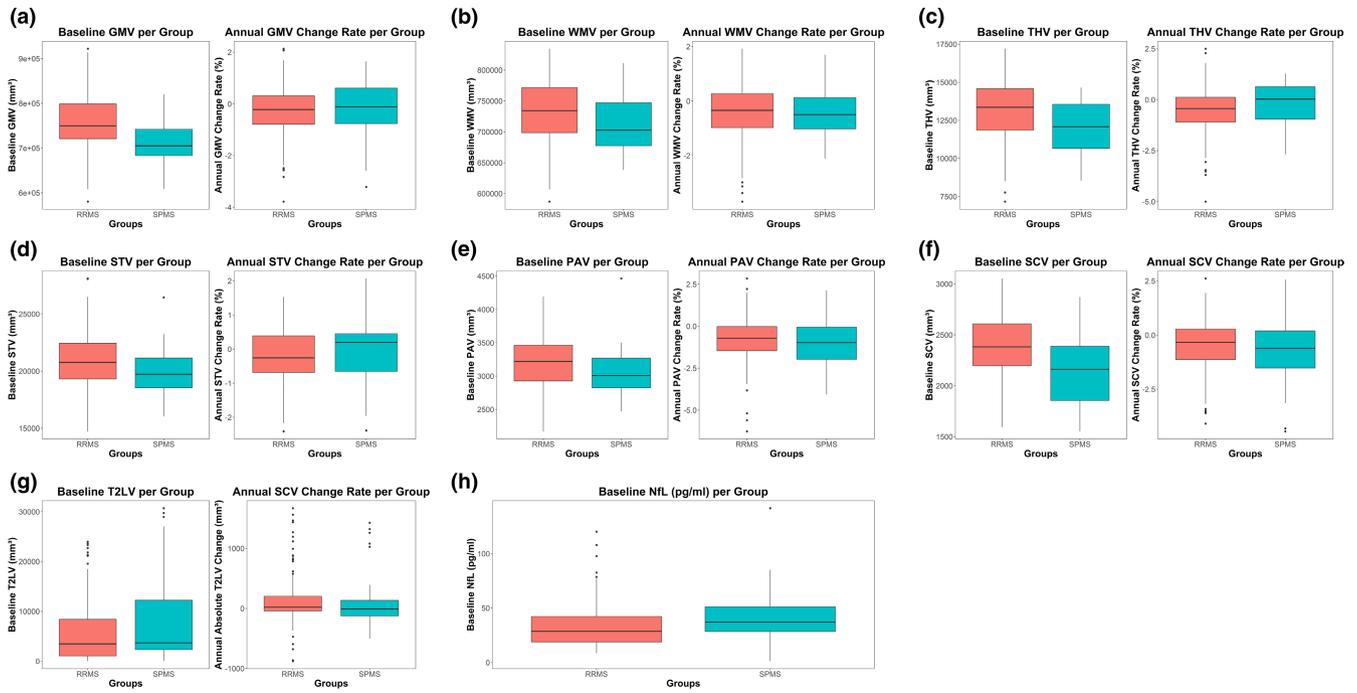


FIGURE 2 Boxplots of baseline and longitudinal magnetic resonance imaging metrics as well as baseline serum neurofilament light chain (NfL) by disease type in relapse-onset multiple sclerosis (MS). Relapsing-remitting (RR) and secondary progressive (SP) MS are depicted with red and turquoise, respectively. (a) GMV, brain gray matter volume, (b) WMV, brain white matter volume, (c) THV, thalamic volume, (d) STV, striatal volume, (e) PAV, pallidal volume, (f) SCV, spinal cord volume, (g) T2LV, T2-weighted lesion volume, (h) NfL. Whiskers correspond to 25th and 75th percentiles [Colour figure can be viewed at wileyonlinelibrary.com]

increase), were significantly associated with higher future ND9HPT worsening over time (mean yearly increase of $\log[\text{ND9HPT}]$ of 0.037 ± 0.011 ; $p = 0.003$). The cross-validation analysis demonstrated highly accurate prediction, with a mean MAPE of 8.7% (95% CI 6.5%–10.9%).

Predictive models of cognitive scores are shown and discussed in Appendix S1 and Tables S3 and S4. SDMT and PASAT scores improved over time in RRMS and were stable in SPMS. Predictive models for SDMT scores had reasonable-to-good predictive accuracy, whereas predictive models for PASAT were inaccurate.

Analyses were also performed for the whole cohort and the final models are presented in Appendix S2 and Table S5. The predictive capabilities of those models were similar to those in the RRMS and SPMS groups.

DISCUSSION

In the present study, we were able to build reliable, robust models capable of accurate predictions of future clinical worsening over time in individual patients (as shown by our leave-one-out fashion cross-validation analysis) while taking respective baseline values into account for each subject. Moreover, there was a dissociation in the prediction of clinical scores between RRMS and SPMS, with different variables predicting future clinical outcomes in these two groups. Finally, our models for the EDSS, 9HPT and T25fwT (only in RRMS) demonstrated high predictive capabilities in our validation

analysis. This study simulates a relatively common real-world clinical scenario of MS patients being regularly assessed with serial MRI as well as clinical examinations, and shows the potential of short-term blood and MRI biomarkers in predicting future disease dynamics.

We were able to build models that not only accurately predict future disease severity but also the dynamics of progression of neurological deficits, as measured by the EDSS. In these models, a dissociation between RRMS and SPMS patients became apparent. On the one hand, the cSC AVCR as well WM lesion load AAVC—reflecting progressive WM injury—arose as predictors of the dynamics of future neurological deterioration in SPMS. With regard to the cSC, our results are in accordance with previous work performed in the same cohort examining concurrent clinical and atrophy changes, which showed that cSC volume loss goes hand in hand with EDSS progression in SPMS patients, with a stronger correlation with EDSS changes compared to RRMS [3]. With regard to WM lesion load, although some previous studies have shown an association between increasing lesion load and disability [25,26], this was not confirmed in other studies [27]. On the other hand, there were no predictors of higher EDSS worsening in RRMS. These between-group differences can be interpreted in a number of ways. First, SPMS patients are older and have sustained a greater toll of neuronal damage over larger periods of time, entailing longer and chronic immune activation, increased oxidative stress-related damage as well as greater loss of trophic support, mitochondrial dysfunction and exhaustion of repair and compensatory mechanisms [28–30]. In addition, in view of the similar cSC ACVR and WM lesion load expansion in both groups,

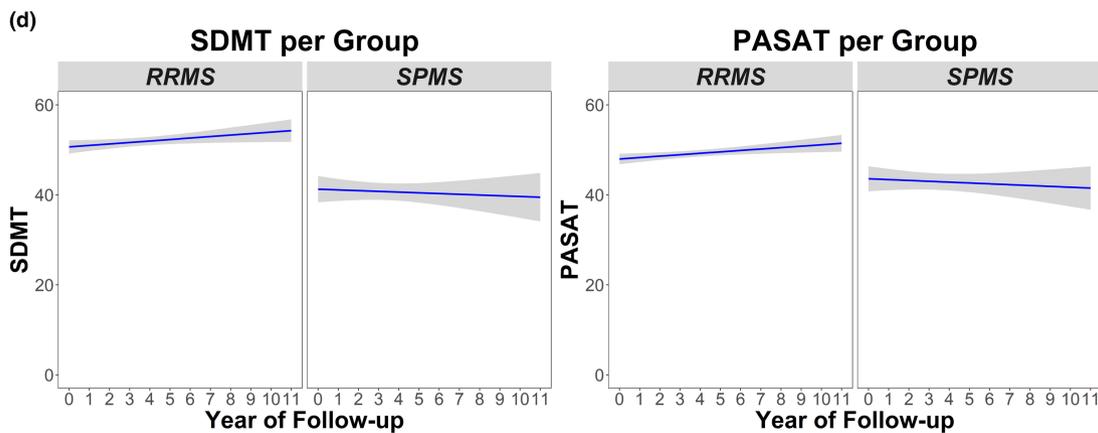
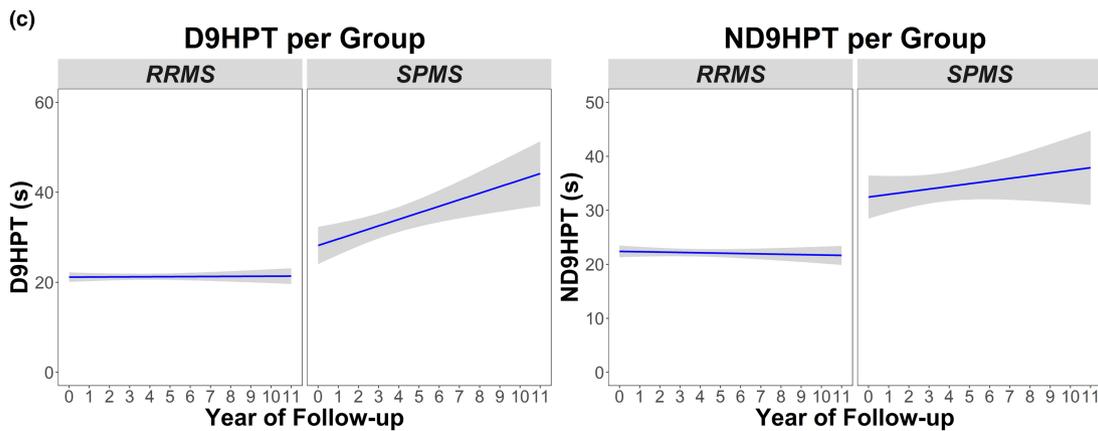
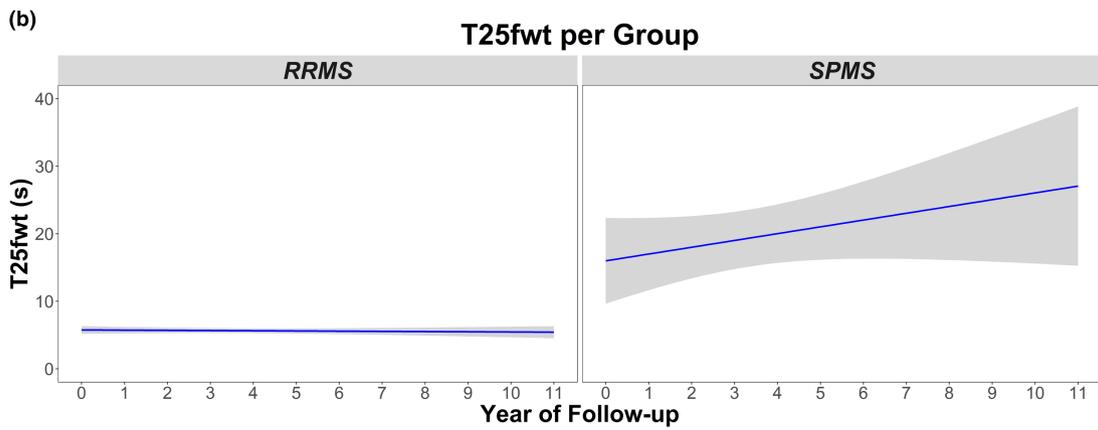
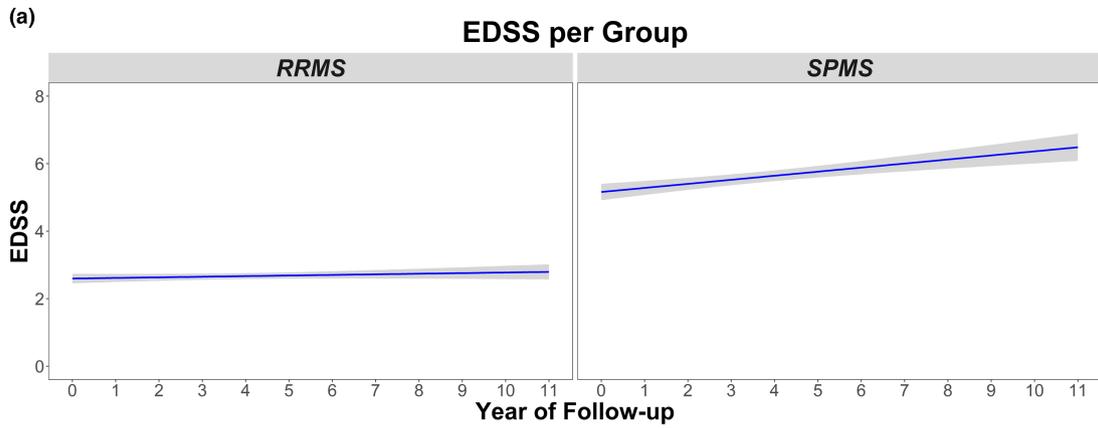


FIGURE 3 Longitudinal trends of the (a) Expanded Disability Status Scale (EDSS), (b) timed 25-foot walk test (T25fwt), (c) dominant hand and non-dominant hand nine-hole peg test (D9HPT and ND9HPT, respectively) as well as (d) Symbol Digit Modalities Test (SDMT) and Paced Auditory Serial Addition (PASAT) are presented over 11 years by disease type. Mean trends are shown in blue lines, 95% confidence intervals are shown in gray [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Final prediction models between clinical scores (Expanded Disability Status Scale, nine-hole peg test and timed 25-foot walk test) and magnetic resonance imaging metrics and baseline neurofilament light chain in relapsing-remitting multiple sclerosis patients

Final models
<p>log(EDSS) ~ Time + Age + Baseline THV + Baseline PAV + Baseline SCV + Annual PAV change rate + (Time Subject) Final model: $R^2_m = 32\%$, $R^2_c = 83\%$, AIC = -216.7 LOOCV: MAE = 0.243 [0.210–0.276], RMSE = 0.269 [0.234–0.303], MAPE = 17.0% [14.7%–19.3%]</p>
<p>1/T25fwt ~ Time + Age + Sex + Baseline GMV + Baseline THV + Baseline PAV + Baseline SCV + annual PAV change rate + Sex:Time + Baseline THV:Time + (Time Subject) Final model: $R^2_m = 32\%$, $R^2_c = 91\%$, AIC = -3663.2 LOOCV: MAE = 0.045 [0.039–0.051], RMSE = 0.048 [0.042–0.054], MAPE = 36.9% [22.3%–51.6%]</p>
<p>log(D9HPT) ~ Time + Sex + Baseline GMV + Contralateral baseline THV + Baseline SCV + Annual T2LV change + Sex:Time + Baseline GMV:Time + (Time Subject) Final model: $R^2_m = 29\%$, $R^2_c = 87\%$, AIC = -846.7 LOOCV: MAE = 0.177 [0.148–0.206], RMSE = 0.193 [0.162–0.224], MAPE = 5.5% [4.8%–6.3%]</p>
<p>log(ND9HPT) ~ Time + Sex + Age + Baseline WMV + Baseline SCV + Annual WMV change rate + Contralateral annual PAV change rate + Annual T2LV changes + Age:Time + Annual T2LV changes:Time + (Time Subject) Final model: $R^2_m = 33\%$, $R^2_c = 92\%$, AIC = -1068.5 LOOCV: MAE = 0.183 [0.149–0.217], RMSE = 0.198 [0.163–0.233], MAPE = 5.7% [4.8%–6.5%]</p>

Note: Analysis was performed with linear mixed effect models with a random intercept and slope denoted as "(Time|Subject)" in our models. The variable "Time" corresponds to the follow-up time measured in years. Variables followed by ":Time" correspond to the interaction between the respective MRI or NFL metric and the time variable, which relates to the effect of the examined metric to the change of the respective clinical score over time (e.g. EDSS change over time). For the purpose of model selection, we used a step-down model-building approach as proposed before, which is based on deletion of effects from the full model using *F*-statistics. In case of significant interaction terms with time, the respective main effects were kept in the model according to the principle of marginality. The initial full model (not shown in this table) included demographics (sex and age), disease duration, medication (injectable, oral, infused), MRI metrics (baseline and annual volume change rates of first three time points) and baseline sNFL values entered as independent variables.

Abbreviations: AIC, akaike information criterion; D9HPT, dominant-hand nine-hole peg test; EDSS, Expanded Disability Status Scale; GMV, cerebral Gray Matter Volume; LOOCV, leave-one-out cross-validation; MAE, mean absolute error; MAPE, mean absolute percentage error; MRI, magnetic resonance imaging; ND9HPT, non-dominant hand nine-hole peg test; NFL, neurofilament light chain; PAV, pallidal volume; R^2_c , conditional *R*-squared; R^2_m , marginal *R*-squared; RMSE, root-mean square error; SCV, spinal cord volume; STV, striatal volume; T25fwt, timed 25-foot walk test; T2LV, T2 lesion volume; THV, thalamic volume; WMV, cerebral white matter volume.

as well as the lower baseline cSC volumes in the SPMS group, it is possible that patients with RRMS still have sufficient reserves of cortical adaptation, remyelination, axonal repair and neuroprotection, which allow them on the one hand to maintain or re-establish the functionality of neuronal tissue and on the other hand to "mask" the produced axonal loss taking place in the cSC and cerebral WM through neuroplasticity occurring at higher cortical centers [31–34]. This allows axonal and myelin damage to be "translated" in a much more straightforward way into clinical deficits once the threshold of neuronal injury and/or repair has been exceeded. Another explanation with regard to cSC atrophy being a significant predictor is that EDSS progression in SPMS is more motor-driven than in RRMS, since the higher EDSS scales, commonly seen in SPMS patients, depend largely on ambulation.

With regard to the prediction of future walking speed, reflecting lower extremity function, we were able to build a model for accurate T25fwt prediction in RRMS. In this model, lower baseline thalamic volumes were associated with a higher T25fwt worsening over time, in line with previous longitudinal studies that point to a longitudinal association between disability and thalamic atrophy progression [2,18]. By contrast, the respective model for SPMS patients demonstrated inaccurate T25fwt predictions. A possible explanation for this might be the considerably higher within- and between-subject variability of these measures in SPMS patients compared with RRMS patients, which can also be visualized in the 95% CI of our T25fwt measures in Figure 3. This may have hampered the construction of accurate predictive models in SPMS patients.

The MRI metrics were also able to produce accurate predictions of future hand dexterity function measurements. In all statistical models, the dominant hand and non-dominant hand were analysed separately, since motor tasks in the non-dominant hand require activations of larger cortical areas in contralateral visuomotor regions including deep GM areas compared to the dominant hand [35]. Indeed, deficits of dominant and non-dominant hand function were shown to be driven by neuronal injury in different CNS structures, with WM (measured either as atrophy or lesion load) and GM injury being associated with non-dominant and dominant hand function worsening, respectively. Widespread interruption of the large network utilized for non-dominant hand functions [35] occurring in the cerebral WM may explain this dissociation. This was evident in the analyses of both RRMS and SPMS patients. In particular, in RRMS, lower baseline GM volume was associated with higher future D9HPT worsening over time, whereas increasing WM injury (in the form of lesion load or WM AVCR) was associated with higher future ND9HPT worsening over time in both RRMS and SPMS patients. In addition, higher cSC atrophy rates, baseline sNFL and baseline striatal volumes were also correlated with ND9HPT worsening over time in SPMS, pointing to a more widespread underlying CNS neurodegenerative pathology

TABLE 4 Final prediction models between clinical scores (Expanded Disability Status Scale, nine-hole peg test and timed 25-foot walk test) and magnetic resonance imaging metrics and baseline neurofilament light chain in secondary progressive multiple sclerosis patients

Final model
<p>log(EDSS) ~ Time + Sex + Disease Duration + Medication + Baseline GMV + Baseline THV + Annual SCV change rate + Baseline NfL + Annual SCV change rate + Annual T2LV change rate + Annual SCV change rate:Time + Annual T2LV change rate:Time + (Time Subject)</p> <p>Final model: $R^2_m = 39\%$, $R^2_c = 95\%$, AIC = -485.6 LOOCV: MAE = 0.191 [0.155-0.227], RMSE = 0.203 [0.164-0.242], MAPE = 11.2% [8.4%-14.1%]</p>
<p>1/T25fwt ~ Time + Sex + Disease Duration + Baseline GMV + Baseline PAV + Baseline T2LV + annual GMV change rate + annual WMV change rate + annual SCV change rate + Sex:Time + Disease Duration:Time + Baseline GMV:Time + Baseline PAV:Time + Baseline T2LV:Time + annual GMV change rate:Time + annual WMV change rate:Time + annual SCV change rate:Time + (Time Subject)</p> <p>Final model: $R^2_m = 15\%$, $R^2_c = 92\%$, AIC = -783.5 LOOCV: MAE = 0.071 [0.054-0.089], RMSE = 0.074 [0.057-0.092], MAPE = 167.9% [71.2%-264.7%]</p>
<p>log(D9HPT) ~ Time + Sex + Medication + Contralateral Baseline THV + Contralateral Baseline STV + Contralateral Baseline PAV + Medication:Time + (Time Subject)</p> <p>Final model: $R^2_m = 22\%$, $R^2_c = 85\%$, AIC = 43.2 LOOCV: MAE = 0.324 [0.250-0.397], RMSE = 0.352 [0.275-0.429], MAPE = 9.2% [7.2%-11.1%]</p>
<p>log(ND9HPT) ~ Time + Sex + Contralateral baseline STV + Baseline NfL + Annual WMV change rate + Annual SCV change rate + Annual T2LV change + Sex:Time + Contralateral baseline STV:Time + Baseline NfL:Time + Annual WMV change rate:Time + Annual SCV change rate:Time + Annual T2LV change:Time + (Time Subject)</p> <p>Final model: $R^2_m = 27\%$, $R^2_c = 90\%$, AIC = -64.6 LOOCV: MAE = 0.310 [0.214-0.407], RMSE = 0.331 [0.234-0.428], MAPE = 8.7% [6.5%-10.9%]</p>

Note: Analysis was performed with linear mixed effect models with a random intercept and slope denoted as "(Time | Subject)" in our models. The variable "Time" corresponds to the follow-up time measured in years. Variables followed by ":Time" correspond to the interaction between the respective MRI or NfL metric and the time variable, which relates to the effect of the examined metric to the change of the respective clinical score over time (e.g. EDSS change over time). For the purpose of model selection, we used a step-down model-building approach as proposed previously, which is based on deletion of effects from the full model using *F*-statistics. In case of significant interaction terms with time, the respective main effects were kept in the model according to the principle of marginality. The initial full model (not shown in this table) included demographics (sex and age), disease duration, medication (injectable, oral, infused), MRI metrics (baseline and annual volume change rates of first three time points) and baseline serum NfL values entered as independent variables.

Abbreviations: AIC, akaike information criterion; D9HPT, dominant hand nine-hole peg test; EDSS, Expanded Disability Status Scale; GMV, cerebral gray matter volume; LOOCV, leave-one-out cross-validation; MAE, mean absolute error; MAPE, mean absolute percentage error; MRI, magnetic resonance imaging; ND9HPT, non-dominant hand nine-hole peg test; NfL, neurofilament light chain; PAV, pallidal volume; R^2_c , conditional *R*-squared; R^2_m , marginal *R*-squared; RMSE, root-mean-square error; SCV, spinal cord volume; STV, striatal volume; T25fwt, timed 25-foot walk test; T2LV, T2 lesion volume; THV, thalamic volume; WMV, cerebral white matter volume.

compared to RRMS. The correlation between high baseline striatal volumes and higher ND9HPT worsening was unexpected and was confirmed, when introducing only baseline striatal volumes into the model (analysis not shown here). Although a series of previous studies have suggested a direct influence of the immune system on the striatum and vice versa [36], in our opinion it is not clear if this association is of a statistical rather than a biological nature.

In addition, an important conclusion drawn from our predictive models is that there is no single one-size-fits-all biomarker that predicts every future clinical outcome. Depending on the clinical outcome of interest, different variables seemed to be crucial predictive factors. Hence, future research in MS patients should rather focus on global assessments of the CNS implementing as many regions of interest rather than isolated atrophy metrics in order to increase the clinical relevance of the findings.

The contribution of *cross-sectional* MRI metrics—next to longitudinal measurements of CNS volume loss—to the prediction of future clinical worsening over time is quite interesting. It could be argued that, apart from the magnitude of ongoing neurodegeneration occurring in the CNS, a decisive predictor of future dynamics of neurological worsening may be brain reserve as shown by higher CNS volumes. According to our results, lower brain reserve is linked not only to higher disability but also to more aggressive progression of neurological deficits. This might be another argument for early treatment in order to retain brain reserve and consequently prevent or mitigate aggressive disease courses.

The present study has a number of limitations. We analysed the follow-up data of an MS cohort in a retrospective manner. Some patients were lost to follow-up during the study, leading to incomplete datasets and potential bias. However, the use of LMER models in our statistical analysis mitigates such issues and is optimal for observational studies. Despite the long observation time of our study, the sample size of SPMS patients was relatively small, which could limit the reproducibility of these results in other SPMS populations. Moreover, the primary goal of this work was to find models that perform well in predicting the dynamics of future clinical outcomes and not to evaluate the utility of individual biomarkers. Hence, we did not account for multicollinearity issues in our analysis because this does not influence the precision of model predictions or the goodness-of-fit statistics. However, multicollinearity hampers the interpretation of the models' individual regression coefficients, and reduces the power of our models to identify independent variables that are statistically significant. Therefore, these values should be treated with caution. In addition, despite the fact that treatment was taken into account in our models, it may well be that the study was underpowered to evaluate effects from different medications, especially since a large number of patients switched treatments during the monitoring time of this study. Future investigations should include greater sample sizes, with patients possibly remaining on single disease-modifying agents. Finally, in this study, MRI scans acquired from a 1.5-T scanner were used, which may have influenced image quality and, as a consequence, the accuracy of our segmentations. Future studies using MRI scans acquired in scanners with a higher magnetic

TABLE 5 Summary of significant predictors for all clinical outcome changes over time in the relapsing-remitting multiple sclerosis, secondary progressive multiple sclerosis groups and the whole cohort

Predictors	Groups	EDSS	T25fwt	D9HPT	ND9HPT	SDMT	PASAT
Baseline GMV	RRMS			X			
	SPMS		X				
	Whole Cohort						
GMV AVCR	RRMS						
	SPMS		X				
	Whole Cohort						
Baseline WMV	RRMS						
	SPMS						
	Whole Cohort			X			
WMV AVCR	RRMS						
	SPMS		X		X	X	
	Whole Cohort						
Baseline THV	RRMS		X				
	SPMS						
	Whole Cohort						
THV AVCR	RRMS						
	SPMS						
	Whole Cohort		X				
Baseline STV	RRMS						
	SPMS				X	X	
	Whole Cohort					X	
STV AVCR	RRMS						
	SPMS						
	Whole Cohort						X
Baseline PAV	RRMS						
	SPMS		X			X	
	Whole Cohort					X	
PAV AVCR	RRMS					X	
	SPMS					X	
	Whole Cohort		X			X	
Baseline SCV	RRMS						
	SPMS						X
	Whole Cohort						
SCV AVCR	RRMS						
	SPMS	X	X		X	X	
	Whole Cohort		X				
Baseline lesion-load	RRMS						
	SPMS		X			X	X
	Whole Cohort						
Lesion-load AAVC	RRMS				X		
	SPMS	X			X		
	Whole Cohort						
Baseline NfL	RRMS						
	SPMS				X		
	Whole Cohort						

Abbreviations: AAVC, absolute annual volume change; AVCR, annual volume change rate; D9HPT, dominant-hand nine-hole peg test; EDSS, Expanded Disability Status Scale; GMV, cerebral gray matter volume; ND9HPT, non-dominant hand nine-hole peg test; NfL, neurofilament light chain; PAV, pallidal volume; RRMS, relapsing-remitting multiple sclerosis; SCV, spinal cord volume; SPMS, secondary progressive multiple sclerosis; STV, striatal volume; T25fwt, timed 25-foot walk test; T2LV, T2 lesion volume; THV, thalamic volume; WMV, cerebral white matter volume.

field (e.g. 3T or 7T) may improve regional segmentation quality due to improved contrast. Nevertheless, the use of a single MRI scanner and a consistent MRI protocol across all scans in our work reduced MRI measurement variability due to technical aspects (e.g. MRI scanner changes, multiple MRI scanners or sequences). We evaluated the prediction of future cognitive outcomes, such as sustained attention and information processing speed deficits using SDMT and PASAT [37,38]. However, these analyses should be interpreted with caution because of two main limitations. Firstly, although cognitive impairment is established in MS [39] and has been shown to increase in long-term longitudinal studies [40], both the SDMT and PASAT in our analysis demonstrated a significant increase over time in RRMS patients, whereas SPMS patients were fairly stable regarding these measures. This can be attributed to a learning effect through repetition, which has also been shown in previous longitudinal studies [41], that may mask “true” cognitive worsening. In addition, PASAT analysis showed very inaccurate predictions of future cognitive performance, which limits the utility of these estimations in clinical practice. It would be important to re-evaluate the prediction of future cognitive impairment using MRI metrics in a different cohort with declining cognitive performance (including broader cognitive domain testing) over time.

In conclusion, the present study demonstrates the capability of short-term MRI metrics to accurately predict future dynamics of neurological disability progression in a large real-world relapse-onset MS cohort. Our results underline the central role of neurodegeneration and provide new insights into the prognostic power of MRI metrics in MS. Due to the long follow-up time (including annual “repeated” measurements), the verification through our cross-validation analysis and the large sample included in this study, we believe that the results are generalizable to other MS populations. Hence, the present work represents a step towards the utilization of structural MRI measurements in patient care.

ACKNOWLEDGMENTS

We are very grateful to all participants and staff involved in the GeneMSA cohort study, in particular Ernst-Wilhelm Radue, MD, and Alain Thoeni. In addition, we would like to thank statistician Andrea Wiencierz, PhD, who provided guidance on the statistical analysis performed by Charidimos Tsagkas in this study. Open access funding provided by Universitat Basel.

CONFLICT OF INTERESTS

Charidimos Tsagkas, Michael Amann, M. Mallar Chakravarty, Cristian Barro have no disclosures. Yvonne Naegelin's employer, the University Hospital Basel received payments for lecturing from Celgene GmbH and Teva Pharma AG that were exclusively used for research support, not related to this study. Athina Papadopoulou has consulted for Teva, received speaker fees from Sanofi-Genzyme and travel support from Bayer AG, Teva, UCB-Pharma AG and Hoffmann La Roche. Her research was/is being supported by the University of Basel, the University Hospital of Basel, the Swiss MS Society, the Swiss National Science Foundation and the “Stiftung

zur Förderung der gastroenterologischen und allgemeinen klinischen Forschung sowie der medizinischen Bildauswertung”. Laura Gaetano is an employee of F. Hoffmann-La Roche Ltd, Basel, Switzerland. Jens Wuerfel is CEO of MIAC AG, Basel, Switzerland, has received speaker honoraria from Bayer, Biogen, Novartis and Teva, has served on advisory boards and received research grants from Biogen and Novartis, and is supported by the German Ministry of Science (BMBF/KKNMS) and the German Ministry of Economy (BMWi). Ludwig Kappos' institution (University Hospital Basel) has received research support and payments that were used exclusively for research support for Ludwig Kappos' activities as principal investigator and member or chair of planning and steering committees or advisory boards in trials sponsored by Actelion, Addex, Almirall, Bayer HealthCare, Celgene, CLC Behring, Genentech, GeNeuro, Genzyme, Merck Serono, Mitsubishi Pharma, Novartis, Octapharma, Ono, Pfizer, Receptos, F. Hoffmann-La Roche, Sanofi-Aventis, Santhera, Siemens, Teva, UCB and XenoPort, licence fees for Neurostatus 4 products, and research grants from the Swiss MS Society, the Swiss National Science Foundation, the European Union, and the Roche Research Foundation. Jens Kuhle reports grants from Biogen, Novartis, Roche, the Swiss MS Society, Sanofi, the University of Basel, the Swiss National Research Foundation, Merck, Celgene and the Progressive MS Alliance, outside the submitted work. The University Hospital Basel, as the employer of Cristina Granziera has received the following fees which were used exclusively for research support: (i) advisory board and consultancy fees from Actelion, Novartis, Genzyme and F. Hoffmann-La Roche; (ii) speaker fees from Biogen and Genzyme-Sanofi; (iii) research support by F. Hoffmann-La Roche Ltd. Before her employment at University Hospital Basel, Cristina Granziera has also received speaker honoraria and travel funding by Novartis. The current (DKD Helios Klinik Wiesbaden) or previous (University Hospital Basel) institutions of Till Sprenger have received payments for speaking or consultation from: Biogen Idec, Eli Lilly, Allergan, Actelion, ATI, Mitsubishi Pharma, Novartis, Genzyme and Teva. Till Sprenger received research grants from the Swiss MS Society, Novartis Pharmaceuticals Switzerland, EFIC-Grünenthal grant, and the Swiss National Science Foundation. Stefano Magon is an employee of F. Hoffmann-La Roche Ltd, Basel, Switzerland. He has received research support from the Swiss MS Society, the Swiss National Science Foundation, the University of Basel and *Stiftung zur Förderung der gastroenterologischen und allgemeinen klinischen Forschung sowie der medizinischen Bildauswertung* University Hospital Basel. Katrin Parmar was supported by the Baasch-Medicus Foundation (2017–2019). Her institution (University Hospital Basel) received speakers' honoraria from Novartis and ExceMED and travel support by Novartis Switzerland.

AUTHOR CONTRIBUTIONS

Charidimos Tsagkas: Conceptualization (lead); Data curation (equal); Formal analysis (lead); Funding acquisition (equal); Investigation (lead); Methodology (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead).

Yvonne Naegelin: Conceptualization (equal); Data curation (equal); Funding acquisition (equal); Investigation (equal); Project administration (lead); Resources (equal); Supervision (equal); Writing – review and editing (equal). **Michael Amann:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – review and editing (equal). **Athina Papadopoulou:** Data curation (equal); Investigation (equal); Project administration (equal); Writing – review and editing (equal). **Christian Barro:** Data curation (equal); Investigation (equal); Methodology (equal); Writing – review and editing (equal). **M. Mallar Chakravarty:** Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – review and editing (equal). **Laura Gaetano:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – review and editing (equal). **Jens Würfel:** Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – review and editing (equal). **Ludwig Kappos:** Conceptualization (equal); Funding acquisition (equal); Resources (equal); Supervision (equal); Writing – review and editing (equal). **Jens Kuhle:** Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing – review and editing (equal). **Cristina Granziera:** Conceptualization (equal); Data curation (equal); Funding acquisition (equal); Investigation (equal); Supervision (equal). **Till Sprenger:** Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review and editing (equal). **Stefano Magon:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Writing – review and editing (equal). **Katrin Parmar:** Conceptualization (equal); Investigation (equal); Methodology (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review and editing (equal).

DATA AVAILABILITY STATEMENT

Anonymized data can be made available upon reasonable request.

ORCID

Charidimos Tsagkas  <https://orcid.org/0000-0002-2586-6316>

Christian Barro  <https://orcid.org/0000-0002-7795-7383>

REFERENCES

- Sloane JA, Mainero C, Kinkel RP. No evidence of disease activity in multiple sclerosis. *JAMA Neurol.* 2015;72(7):835-836.
- Eshaghi A, Prados F, Brownlee WJ, et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. *Ann Neurol.* 2018;83(2):210-222.
- Tsagkas C, Magon S, Gaetano L, et al. Spinal cord volume loss: a marker of disease progression in multiple sclerosis. *Neurology.* 2018;91(4):e349-e358.
- Barro C, Benkert P, Disanto G, et al. Serum neurofilament as a predictor of disease worsening and brain and spinal cord atrophy in multiple sclerosis. *Brain.* 2018;141(8):2382-2391.
- Baranzini SE, Wang J, Gibson RA, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet.* 2009;18(4):767-778.
- Weier K, Mazraeh J, Naegelin Y, et al. Biplanar MRI for the assessment of the spinal cord in multiple sclerosis. *Mult Scler J.* 2012;18(11):1560-1569.
- Tsagkas C, Magon S, Gaetano L, et al. Preferential spinal cord volume loss in primary progressive multiple sclerosis. *Mult Scler.* 2018;25(7):947-957.
- McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann Neurol.* 2001;50(1):121-127.
- Thebault S, Abdoli M, Fereshtehnejad S-M, Tessier D, Tabard-Cossa V, Freedman MS. Serum neurofilament light chain predicts long term clinical outcomes in multiple sclerosis. *Sci Rep.* 2020;10(1):10381.
- Magon S, Gaetano L, Chakravarty MM, et al. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neurosci.* 2014;15(1):106.
- Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage.* 2002;17(1):479-489.
- Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage.* 2004;23(Supplement 1):S208-S219.
- Hirai T, Jones EG. A new parcellation of the human thalamus on the basis of histochemical staining. *Brain Res Brain Res Rev.* 1989;14(1):1-34.
- Chakravarty MM, Bertrand G, Hodge CP, Sadikot AF, Collins DL. The creation of a brain atlas for image guided neurosurgery using serial histological data. *NeuroImage.* 2006;30(2):359-376.
- Chakravarty MM, Steadman P, van Eede MC, et al. Performing label-fusion-based segmentation using multiple automatically generated templates. *Hum Brain Mapp.* 2013;34(10):2635-2654.
- Pipitone J, Park MTM, Winterburn J, et al. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. *NeuroImage.* 2014;110(1):494-512.
- Magon S, Chakravarty M, Amann M, et al. Label-fusion-segmentation and deformation-based shape analysis of deep gray matter in multiple sclerosis: the impact of thalamic subnuclei on disability. *Hum Brain Mapp.* 2014;35(8):4193-4203.
- Magon S, Tsagkas C, Gaetano L, et al. Volume loss in the deep gray matter and thalamic subnuclei: a longitudinal study on disability progression in multiple sclerosis. *J Neurol.* 2020;267(5):1536-1546.
- Sdika M, Pelletier D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Hum Brain Mapp.* 2009;30(4):1060-1067.
- Magon S, Gaetano L, Chakravarty MM, et al. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neurosci.* 2014;15(1):106.
- Amann M, Pezold S, Naegelin Y, et al. Reliable volumetry of the cervical spinal cord in MS patient follow-up data with cord image analyzer (Cordial). *J Neurol.* 2016;263(7):1364-1374.
- Diggle PJ, Heagerty P, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data.* OUP; 2002:396.
- Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM. *Mixed Effects Models and Extensions in Ecology with R* [Internet]. Springer-Verlag; 2009 [cited 2020 Aug 8]. (Statistics for Biology and Health). Available from: <https://www.springer.com/gp/book/9780387874579>
- Lewis CD. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting.* Butterworth Scientific; 1982:143.
- Kearney H, Rocca MA, Valsasina P, et al. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. *Mult Scler.* 2014;20(1):72-80.

26. Ammitzbøll C, Dyrby TB, Lyksborg M, et al. Disability in progressive MS is associated with T2 lesion changes. *Mult Scler Relat Disord*. 2018;20:73-77.
27. Popescu V, Agosta F, Hulst HE, et al. Brain atrophy and lesion load predict long term disability in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2013;84(10):1082-1091.
28. Bjartmar C, Wujek JR, Trapp BD. Axonal loss in the pathology of MS: consequences for understanding the progressive phase of the disease. *J Neurol Sci*. 2003;206(2):165-171.
29. Larochelle C, Uphaus T, Prat A, Zipp F. Secondary progression in multiple sclerosis: neuronal exhaustion or distinct pathology? *Trends Neurosci*. 2016;39(5):325-339. Oct [cited 2016 Mar 30]; Available from: <http://www.sciencedirect.com/science/article/pii/S0166223616000254>
30. de Barcelos IP, Troxell RM, Graves JS. Mitochondrial dysfunction and multiple sclerosis. *Biology*. 2019;8(2):37.
31. Goldschmidt T, Antel J, König FB, Brück W, Kuhlmann T. Remyelination capacity of the MS brain decreases with disease chronicity. *Neurology*. 2009;72(22):1914-1921.
32. Tomassini V, Matthews PM, Thompson AJ, et al. Neuroplasticity and functional recovery in multiple sclerosis. *Nat Rev Neurol*. 2012;8(11):635-646.
33. Ksiazek-Winiarek DJ, Szpakowski P, Glabinski A. Neural plasticity in multiple sclerosis: the functional and molecular background. *Neural Plasticity*. 2015;2015:1-11.
34. Nakahara J. Remyelination in multiple sclerosis: pathology and treatment strategies. *Clin Exp Neuroimmunol*. 2017;8(S1):40-46.
35. Alahmadi AAS, Pardini M, Samson RS, et al. Differential involvement of cortical and cerebellar areas using dominant and nondominant hands: an fMRI study: effects of forces on motor task using both hands. *Hum Brain Mapp*. 2015;36(12):5079-5100.
36. Mancini A, Ghiglieri V, Parnetti L, Calabresi P, Di Filippo M. Neuro-immune cross-talk in the striatum: from basal ganglia physiology to circuit dysfunction. *Front Immunol*. 2021;12:644294. [cited 2021 Jun 15];12. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.644294/full>
37. Tombaugh TN. A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Arch Clin Neuropsychol*. 2006;21(1):53-76.
38. Benedict RH, DeLuca J, Phillips G, et al. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler*. 2017;23(5):721-733.
39. Sumowski JF, Benedict R, Enzinger C, et al. Cognition in multiple sclerosis: State of the field and priorities for the future. *Neurology*. 2018;90(6):278-288.
40. Amato MP, Ponziani G, Siracusa G, Sorbi S. Cognitive dysfunction in early-onset multiple sclerosis: a reappraisal after 10 years. *Arch Neurol*. 2001;58(10):1602-1606.
41. Conradsson D, Ytterberg C, von Koch L, Johansson S. Changes in disability in people with multiple sclerosis: a 10-year prospective study. *J Neurol*. 2018;265(1):119-126.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Tsagakas C, Naegelin Y, Amann M, et al. Central nervous system atrophy predicts future dynamics of disability progression in a real-world multiple sclerosis cohort. *Eur J Neurol*. 2021;28:4153-4166. <https://doi.org/10.1111/ene.15098>