Supplementary material

Index

eTable 1. Overview of MRI protocols for T1-weighted images

	T1-weighted acquisition					
Center	Lausanne	Geneva		Basel		Lugano
Number of subjects	12	49		354		30
Scanner vendor	Siemens	Siemens	Siemens	Siemens	Siemens	Siemens
Scanner model	Skyra	Aera	Skyra	Avanto	Skyra_fit	Skyra
Magnetic field	3 Tesla	1.5 Tesla	3 Tesla	1.5 Tesla	3 Tesla	3 Tesla
TR (ms)	2300	2200	2000	2700	2300	2300
TE (ms)	2.9	2.67	2.01	5.03	3.02	2.98
Inversion time (ms)	900	900	1100	950	900	900
Matrix size	256x240	256x256	256x256	256x256	256x256	256x256
FOV	256x240	256x256	256x256	256x256	256x256	256x256
Resolution (mm)	1x1x1	1x1x1	1x1x1	1x1x1	1x1x1	1x1x1
Flip angle	9	8	9	8	9	9
Acquisition type	3D	3D	3D	3D	3D	3D

Abbreviations: TR, repetition time; TE, echo time; TI, inversion time; FOV, field of view.

eTable 2. Overview of MRI protocols for 3D-EPI images

	3D-EPI acquisition			
Center	Lausanne	Geneva	Basel	Lugano
Nr of patients	12	49	354	30
Scanner vendor	Siemens	Siemens	Siemens	Siemens
Scanner model	Skyra	Skyra	Skyra_fit	Skyra
Magnetic field	3 Tesla	3 Tesla	3 Tesla	3 Tesla
TR (ms)	64	65	65	65
TE (ms)	35	36	36	36
Inversion time (ms)	/	/	/	/
Matrix size	384x312	352x286	384x312	352x286
FOV	250x203	250x203	250x203	250x203
Resolution (mm)	0.65x0.65x0.65	0.71x0.71x0.71	0.65x0.65x0.85	0.71x0.71x0.71
Flip angle	10	10	10	10
Acquisition type	3D	3D	3D	3D

Abbreviations: 3D-EPI, three-dimensional echo planar imaging; TR, repetition time; TE, echo time; TI, inversion time; FOV, field of view.

eMethods 1

In the statistical analyses disease-modifying therapies (DMTs) class was defined as follows: group 1 (platform DMTs), which included interferon-beta preparations, and glatiramer-acetate; group 2 (oral DMTs), which included teriflunomide, dimethyl fumarate, fingolimod, and cladribine; group 3 (monoclonal antibodies), which included natalizumab, rituximab, and ocrelizumab. Untreated patients were considered as a separate group.

In the statistical analyses, MRI protocol was defined by the combination of center, and MRI scanner used to acquire the images.

Relapses were defined as new, worsening or recurrent neurological symptoms that lasted for at least 24 hours without fever, infection, or adverse reaction to a prescribed medication and that were preceded by a stable or improving neurological status of at least 30 days.

eAppendix 1 – Additional analyses on the predictive value of baseline MRI measurements on time-to-PIRA

1. Univariable Cox-regression analyses

Results of the univariable Cox proportional-hazards models exploring the association between i) baseline BPF, ii) baseline C2-C3 CSA, iii) baseline thalamic fraction, and iv) baseline cortical fraction, and time-to-PIRA are reported in **eTable 10**.

eTable 3. Predictive value of baseline MRI volumetric measurements in univariable Cox-regression models

	HR (95% CI)	p-value
Baseline BPF	0.573 (0.452; 0.727)	**<0.0001**
Baseline C2-C3 CSA	0.623 (0.480; 0.809)	**<0.0001**
Baseline thalamic fraction	0.723 (0.563; 0.927)	**0.0106**
Baseline cortical fraction	0.655 (0.576; 0.757)	**<0.0001**

Abbreviations: BPF, brain parenchymal fraction; CSA, cross-sectional area; HR, hazard ratio; CI, confidence interval.

2. Predictive value of baseline T2-hyperintense lesion volume

We explored the association between baseline T2-hyperintense lesion volume (after logarithmic transformation, log-T2LV) and time-to-PIRA in both univariable and multivariable Cox proportional-hazards models. In the multivariable model, age, sex, disease duration, DMT class, and MRI protocol were included as covariates.
Baseline log-T2LV was a significant predictor of time-to-PIRA in both the univariable [HR: 1.486 (95% CI: 1.207; 1.830; p=0.0002)], and the multivariable models [HR: 1.304 (95% CI: 1.036; 1.642); p=0.024].

When including log-T2LV as an additional covariate in the multivariable Cox regression models exploring the predictive value of i) baseline C2-C3 CSA, ii) baseline BPF, iii) baseline thalamic fraction, and iv) baseline cortical fraction, the results remained substantially unchanged compared to the original analyses. Specifically:

- in the model with baseline C2-C3 CSA: baseline C2-C3 CSA was associated with time-to-PIRA [HR: 0.634 (95% CI: 0.463; 0.869); p=0.005], while baseline log-T2LV was not [HR: 1.253 (95% CI: 0.989; 1.589); p=0.062];
- in the model with baseline BPF: baseline BPF was associated with time-to-PIRA [HR: 0.651 (95% CI: 0.449; 0.945); p=0.024], while baseline log-T2LV was not [HR: 1.139 (95% CI: 0.884; 1.469); p=0.31];
- in the model with baseline thalamic fraction: baseline log-T2LV was associated with time-to-PIRA [HR: 1.366 (95% CI: 0.022; 0.827); p=0.035], while baseline thalamic fraction was not [HR: 1.109 (95% CI: 0.742; 1.656); p=0.62];
- in the model with baseline cortical fraction: baseline cortical fraction was associated with time-to-PIRA [HR: 0.702 (95% CI: 0.552; 0.893); p=0.004], while baseline log-T2LV was not [HR: 1.214 (95% CI: 0.962; 1.532); p=0.10].

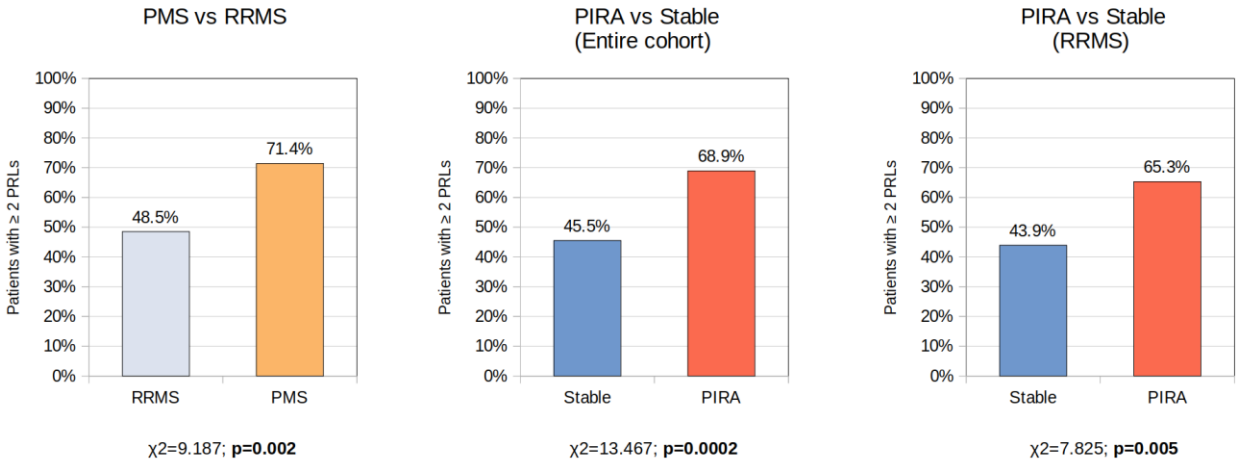
3. Multivariable model including all baseline MRI volumetric measurements

To assess the independent value of all baseline MRI volumetric measurements in predicting time-to-PIRA we used a multivariable Cox proportional-hazards model including as covariates: i) baseline C2-C3 CSA, ii) baseline BPF, iii) baseline thalamic fraction, iv) baseline cortical fraction, and v) baseline log-T2LV – in addition to age, sex, disease duration, DMT class, and MRI protocol. The model did not show severe multicollinearity [variance inflation factors (VIF) < 5].
In such a model, while baseline C2-C3 CSA [HR: 0.648 (95% CI: 0.466; 0.902); p=0.010], and baseline cortical fraction [HR: 0.746 (95% CI: 0.566; 0.983); p=0.037] remained significantly associated with time-to-PIRA, baseline BPF [HR: 0.677 (95% CI: 0.432; 1.060); p=0.088], baseline thalamic fraction [HR: 1.542 (95% CI: 0.943; 2.520); p=0.084], and baseline log-T2LV [HR: 1.243 (95% CI: 0.928; 1.666); p=0.145] were not.

eAppendix 2 – **Additional analyses on between-group comparison in PRL burden**

Between-group comparisons in PRL burden were performed also after dichotomizing patients based on a cut-off of 2 PRLs, as previously proposed by Maggi *et al.*[1] Such threshold has been shown to identify patients exhibiting more aggressive clinical course, and increased levels of serum neurofilament light chain – a marker of neuroaxonal damage.[1]
Between-group comparisons in the proportion of subject with ≥ 2 PRLs were performed with Pearson's chi-square test.
Results are displayed in eFigure 3.

eFigure 1. Group comparisons in the proportion of patients with at least 2 PRLs



Abbreviations: PMS, progressive multiple sclerosis; RRMS, relapsing-remitting multiple sclerosis; PIRA, progression independent of relapse activity; PRLs: paramagnetic rim lesions.

eAppendix 3 – Multivariable logistic regression model including spinal cord CSA and PRL count

To assess the relative strength of association between the occurrence of PIRA during follow-up and i) C2-C3 CSA, and ii) PRL burden we used a multivariable logistic regression model. PIRA status was used as a binary dependent variable, and C2-C3 CSA and PRL count measured at the end of follow-up (when both variables were concomitantly available) were included as explanatory variables, adjusting for age, sex, disease duration, log-T2LV, and DMTs class. In such a model, both C2-C3 CSA [OR: 0.952 (95% CI: 0.915; 0.988); p=0.012] and PRL count [OR: 1.081 (95% CI: 1.032; 1.134); p=0.001] were independently associated with the occurrence of PIRA during the 4-year follow-up, while log-T2LV was not [OR: 0.741 (95% CI: 0.504; 1.076); p=0.12].

eAppendix 4 – Spinal cord atrophy and paramagnetic rim lesions in patients with RAW

1. Baseline C2-C3 cross-sectional area

In a multivariable linear regression model with C2-C3 CSA as dependent variable, and patient group as independent variable, adjusting for age, sex, disease duration, total intracranial volume (TIV), disease-modifying therapies (DMTs) class, and MRI protocol, patients that during follow-up presented RAW had lower baseline C2-C3 CSA compared to patients that did not present confirmed disability accumulation (CDA) [b=-5.581 (95% CI: -9.378; -1.782); p=0.004]. No significant difference in baseline C2-C3 CSA was measured between patients presenting RAW and patients presenting PIRA during follow-up [b=-2.339 (95% CI: -6.507; 1.830); p=0.27].

2. Longitudinal change in C2-C3 cross-sectional area

Between-group differences were modeled using an interaction term between patient group and time in a linear mixed-effect model using the C2-C3 CSA at each given time point as dependent variable. Models included subjects and MRI protocol as random intercepts, and a random slope on time. As fixed-effect covariates we considered time, age at baseline, sex, disease duration at baseline, TIV, DMTs class at baseline, as well as the interactions between age at baseline and sex with time.
During observation, there was not significant difference in the rate of spinal cord atrophy between patients with RAW and patients without CDA [MD-APC: 1.28 (95%CI: -0.05; 2.64); p=0.07]. Patients exhibiting PIRA had accelerated spinal cord atrophy compared to patients with RAW [MD-APC: -2.71 (95% CI: -1.19; -4.26); p=0.0006].

3. PRL count

In an univariable negative binomial regression model, no significant differences in PRL count between patient with RAW and patients without CDA [IRR: 1.851 (95% CI: 0.920; 4.365); p=0.12], as well as between patients with RAW and patients with PIRA [IRR: 0.914 (95% CI: 0.420; 2.264); p=0.83] were measured.

eAppendix 5 – Sensitivity analyses on MRI protocol

1. Heterogeneity in magnetic field strength

The inclusion of MRI images obtained with different magnetic field strengths may constitute a bias in the estimation of longitudinal rates of spinal cord atrophy, which we tried to overcome by systematically accounting for it in the statistical analyses. Aiming at further corroborating our analyses, we also assessed the consistency of the results by repeating the analyses after the exclusion of the scans acquired with 1.5 Tesla scanners. Specifically, 165 MRI scans (and 14 patients) were excluded.

The exclusion of scans performed at 1.5 Tesla did not significantly modify the results. Specifically:

- the mean difference in annual percentage change (MD-APC) of C2-C3 CSA between PMS and RRMS was -1.35 (95% CI: -2.35; -0.36); p=0.008;
- the MD-APC of C2-C3 CSA between patients with PIRA and patients without CDA was -1.38 (95% CI: -2.17; -0.60); p=0.0006;
- the MD-APC of C2-C3 CSA between patients with PIRA and patients without CDA, when restricting the analysis to patients with an RRMS disease course, was -1.06 (95% CI: -1.96; -0.16); p=0.022.

2. Heterogeneity in MRI scanners

The potential confounding effect of the inclusion of scans obtained with heterogeneous MRI scanners across centers has been address by accounting for it in all statistical analyses. Notably, the inter-center differences in MRI images were also mitigated by the use across the SMSC centers of MRI protocols with homogenized signal-to-noise ratio. Nevertheless, we also assessed the consistency of our results by repeating the analyses in the subset of patients belonging to the most represented center (Basel, which included 354 patients). In this subset, the results were not significantly altered. Specifically:

- the MD-APC of C2-C3 CSA between PMS and RRMS was -1.59 (95% CI: -2.56; -0.61); p=0.001;
- the MD-APC of C2-C3 CSA between patients with PIRA and patients without CDA was -1.37 (95% CI: -2.14; -0.59); p=0.0007;
- the MD-APC of C2-C3 CSA between patients with PIRA and patients without CDA, when restricting the analysis to patients with an RRMS disease course, was -1.09 (95% CI: -1.97; -0.22); p=0.016.

Between-group comparisons in PRL burden were also repeated in a subset of patients belonging to a single center (Basel, which was the most represented center). In this subgroup of patients, the results obtained in the entire cohort were confirmed:
• Compared to patients with RRMS, patients with PMS had increased PRL count [IRR: 2.049 (95% CI: 1.261-3.539); p=0.006]
• Compared to patients without CDA, patients with PIRA had increased PRL count [IRR: 1.964 (95% CI: 1.297-3.073); p=0.002]
• In the subgroup of patients with RRMS, patients with PIRA had increased PRL count compared to patients without CDA [IRR: 1.767 (95% CI: 1.081-3.049); p=0.031].

eAppendix 6 - Sensitivity analyses excluding patients with PPMS

In the study cohort 14 subjects (3.1%) had a diagnosis of PPMS, of whom 8 presented PIRA episodes (10.8% of patients with PIRA). To exclude that the observed results in our cohort were significantly driven by the subgroup of patients with PPMS, we performed a sensitivity analysis excluding such patients.

eTable 4 Comparison in baseline C2-C3 CSA between PIRA and stable patients: entire cohort vs excluding patients with PPMS

	Entire cohort		**Excluding PPMS**	
	b (95% CI)	p-value	*b (95% CI)*	p-value
C2-C3 CSA: PIRA vs Stable	-3.188 (-5.312; -1.065)	**0.003**	-2.808 (-5.037; -0.580)	**0.012**

Abbreviations: PPMS, primary progressive multiple sclerosis; CSA, cross-sectional area; PIRA, progression independent of relapse activity; b, regression coefficient; CI, confidence interval.

eTable 5 Comparison in spinal cord atrophy between PIRA and stable patients: entire cohort vs excluding patients with PPMS

	Entire cohort		Excluding PPMS	
	MD-APC (95% CI)	p-value	*MD-APC (95% CI)*	p-value
C2-C3 CSA change: PIRA vs Stable	-1.39 (-2.18; -0.59)	**0.0008**	-1.45 (-2.30; -0.59)	**0.001**

Abbreviations: PPMS, primary progressive multiple sclerosis; CSA, cross-sectional area; PIRA, progression independent of relapse activity; MD-APC, mean difference in annual percentage change; CI, confidence interval.

eTable 6 Baseline MRI measurements for prediction of time-to-PIRA: entire cohort vs excluding patients with PPMS

	Entire cohort		Excluding PPMS	
	HR (95% CI)	p-value	*HR (95% CI)*	p-value
Baseline C2-C3 CSA	0.61 (0.45; 0.83)	**0.001**	0.62 (0.44; 0.86)	**0.005**
Baseline BPF CSA	0.59 (0.43; 0.82)	**0.002**	0.61 (0.43; 0.86)	**0.005**
Baseline thalamic fraction	0.83 (0.60; 1.13)	0.24	0.82 (0.58; 1.15)	0.25
Baseline cortical fraction	0.68 (0.54; 0.85)	**0.0008**	0.61 (0.42; 0.87)	**0.007**

Abbreviations: PPMS, primary progressive multiple sclerosis; CSA, cross-sectional area; BPF, brain parenchymal fraction; HR, hazard-ratio; CI, confidence interval.

eTable 7 Comparison in PRL count between PIRA and stable patients: entire cohort vs excluding patients with PPMS

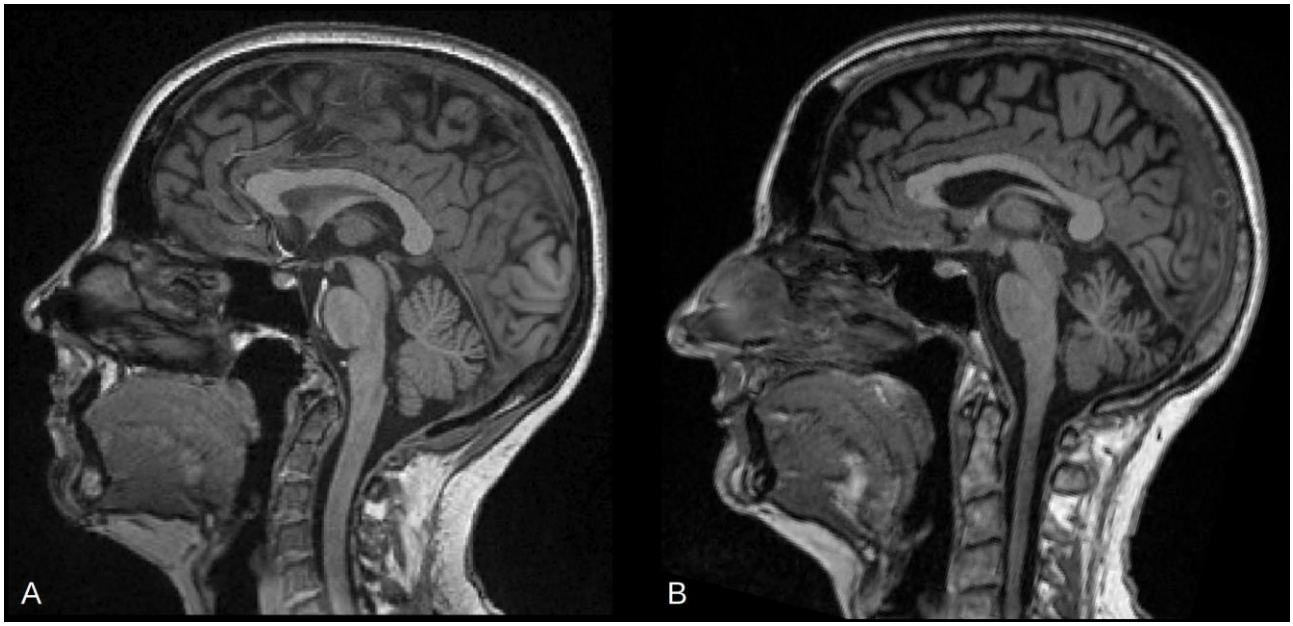
	Entire cohort		Excluding PPMS	
	IRR (95% CI)	p-value	*IRR (95% CI)*	p-value
C2-C3 CSA change: PIRA vs Stable	1.95 (1.34; 2.92)	**0.0008**	1.93 (1.29; 3.01)	**0.002**

Abbreviations: PPMS, primary progressive multiple sclerosis; CSA, cross-sectional area; PIRA, progression independent of relapse activity; IRR, incidence rate ratio; CI, confidence interval.

eAppendix 7 – Cross-validation

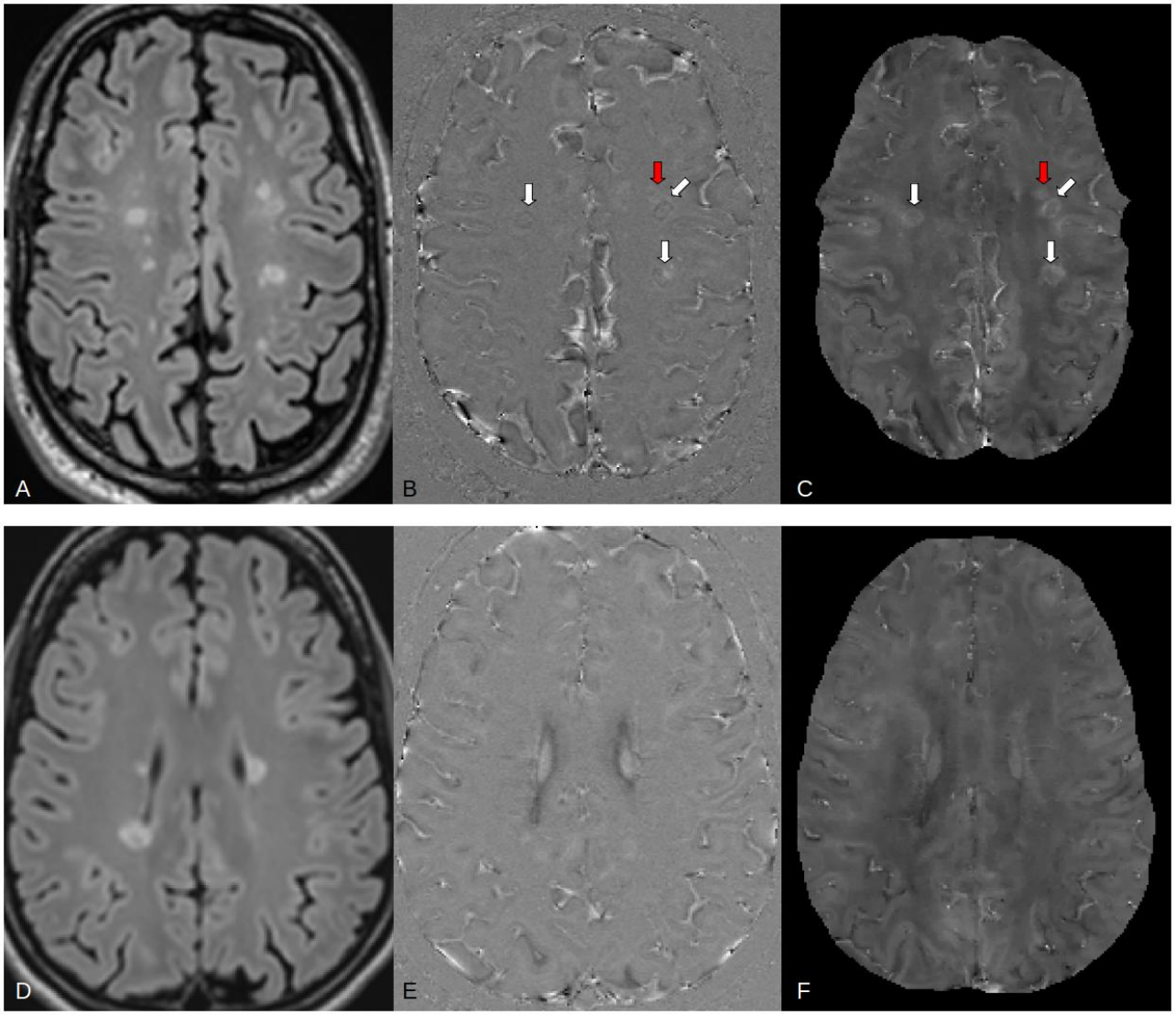
Cross-validation was used to assess the generalization ability of the models used in our study. Specifically, a 10-fold cross-validation was applied to: i) Cox regression models used to predict time-to-PIRA, and ii) linear mixed-effect model investigating the rates of spinal cord atrophy, using the R packages 'rms' (https://cran.r-project.org/web/packages/rms/rms.pdf) and 'cvms' (https://cran.r-project.org/web/packages/cvms/cvms.pdf), respectively. We estimated the model performance with Somers' D_{xy} rank correlation for Cox regression models, and with the root mean square error (RMSE) for the linear-mixed effect model.
For Cox regression, the D_{xy} in the original model vs in cross-validation was: i) 0.441 vs 0.403 for the model including baseline C2-C3 CSA, ii) 0.486 vs 0.431 for the model including baseline BPF, iii) 0.427 vs 0.325 for the model including baseline thalamic fraction, and iv) 0.468 vs 0.423 for the model including baseline cortical fraction.
In the linear-mixed effect model, the RMSE was 0.0178 in the entire cohort, and 0.0593 in cross-validation.

eFigure 2 – Example of MRI images from patients without and with spinal cord atrophy



Legend: A. Example of a preserved cervical spinal cord in a 48 years old woman with RRMS and clinical stability during follow-up (T1-weighted MPRAGE (native) image, sagittal plane); B. Example of pronounced cervical spinal cord atrophy in a 46 years old woman with RRMS who experienced PIRA during follow-up (T1-weighted MPRAGE (native) image, sagittal plane).

eFigure 3 – Example of MRI images from patients with and without paramagnetic rim lesions



Legend: Example of FLAIR-hyperintense white matter lesions (A) exhibiting a rim of paramagnetic signal in both unwrapped phase (B) and quantitative susceptibility mapping (C) in a 34 years old man with RRMS who experienced PIRA during follow-up. Example of FLAIR-hyperintense white matter lesions (D) without a paramagnetic rim on unwrapped phase (E) and quantitative susceptibility mapping (F) in a 37 year old woman with RRMS and clinical stability during follow-up.
White arrows indicate paramagnetic rim lesions; the lesion indicated by the red arrow showed a paramagnetic rim that is only partially visible in this slice.
Panels A, D: axial native FLAIR;
Panels B, E: axial unwrapped filtered phase;
Panels C, F: axial quantitative susceptibility mapping (QSM).

eTable 8. Clinical characteristics of patients across SMSC centers

	Clinical characteristics

Center	Lausanne	Geneva	Basel	Lugano
Nr of patients	12	49	354	30
Female, No. (%)	10 (83.3)	32 (65.3)	229 (64.7)	18 (60.0)
Age at baseline, mean (SD), years	46.9 (11.2)	39.5 (9.0)	45.9 (11.6)	43.2 (10.5)
Disease duration at baseline, median (IQR), years	9.0 (6.0-14.5)	5.4 (3.2-9.3)	11.4 (6.6-19.5)	10.8 (6.9-14.2)
EDSS at baseline, median (IQR)	1.5 (1.5-2.125)	1.5 (1.0-2.5)	2.5 (1.5-3.5)	2.5 (2.0-3.5)
Disease course at baseline - RRMS, No. (%) - SPMS, No. (%) - PPMS, No. (%)	 12 (100) 0 (0) 0 (0)	 42 (85.7) 5 (10.2) 2 (4.1)	 312 (88.1) 31 (8.8) 11 (3.1)	 29 (96.7) 0 (0) 1 (3.3)
Subjects on DMTs, No. (%) - Platform, No. - Oral, No. - Monoclonal antibodies, No.	10 (83.3) 0 6 4	43 (87.8) 3 25 15	301 (85.0) 35 203 63	24 (80.0) 3 7 14
Clinical evolution during follow-up - Stable, No. - PIRA, No. - RAW, No.	 12 0 0	 41 4 4	 274 67 13	 27 3 0

Abbreviations: SD, standard deviation; IQR, interquartile range; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary progressive multiple sclerosis; PPMS, primary progressive multiple sclerosis; DMTs, disease-modifying therapies; PIRA, progression independent of relapse activity; RAW, relapse-associated worsening.

eTable 9. Clinical characteristics of patients by disease phenotype

	Clinical characteristics		
Disease phenotype	RRMS	SPMS	PPMS
Nr of patients	395	36	14
Female, No. (%)	264 (66.8)	18 (50.0)	7 (50.0)
Age at baseline, mean (SD), years	43.8 (11.0)	53.5 (10.5)	57.0 (8.4)
Disease duration at baseline, median (IQR), years	9.4 (5.5-16.3)	20.9 (15.3-30.9)	9.4 (7.7-17.0)
EDSS at baseline, median (IQR)	2.0 (1.5-3.0)	5.5 (4.5-6.125)	4.250 (3.625-6.0)
Subjects on DMTs, No. (%) - Platform, No. - Oral, No. - Monoclonal antibodies, No.	345 (87.3) 36 231 78	29 (80.6) 5 10 14	4 (28.6) 0 0 4
Clinical evolution during follow-up - Stable, No. - PIRA, No. - RAW, No.	 331 49 15	 17 17 2	 6 8 0

Abbreviations: SD, standard deviation; IQR, interquartile range; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary progressive multiple sclerosis; PPMS, primary progressive multiple sclerosis; DMTs, disease-modifying therapies; PIRA, progression independent of relapse activity; RAW, relapse-associated worsening.

eTable 10. Additional clinical data

	Cohort (n=445)	PIRA (n=74)	RAW (n=17)	Patients without CDA (n=354)
Relapses during follow-up, No.	107	10	22	75
Patients with relapse activity during follow-up, No. (%)	84 (18.9)	8 (10.8)	17 (100)	59 (16.7)
Patients treated with steroids during follow-up, No. (%)	54 (12.1)	6 (8.1)	11 (64.7)	37 (10.5)
Patients switching DMTs during follow-up, No. (%)	182 (40.9)	38 (51.4)	13 (76.5)	131 (37.0)
DMTs switches during follow-up, No. - same class, No. (%) - to higher class, No. (%) - to lower class, No. (%)	217 75 (34.6) 115 (53.0) 27 (12.4)	44 14 (31.8) 25 (56.8) 5 (11.4)	19 5 (26.3) 12 (63.2) 2 (10.5)	154 56 (36.4) 78 (50.6) 20 (13.0)

eReferences

1. Maggi P, Kuhle J, Schädelin S, van der Meer F, Weigel M, Galbusera R, et al. Chronic White Matter Inflammation and Serum Neurofilament Levels in Multiple Sclerosis. Neurology [Internet]. 2021 Aug 8 [cited 2022 Sep 12];97(6):e543–53. Available from: /pmc/articles/PMC8424501/