Supplementary Online Content

Cantó E, Barro C, Zhao C, et al. Association between serum neurofilament light chain levels and long-term disease course among patients with multiple sclerosis followed up for 12 years. *JAMA Neurol*. Published online August 12, 2019. doi:10.1001/jamaneurol.2019.2137

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Treatments

Considering the complex individual clinical trajectories in real-world cohorts, to prevent losing power due to multiple stratifications, we dichotomized treatment histories as has been done in previous studies [1]. Therapies referred to as "platform therapy" included self-injectables (interferon (IFN) beta-1b, IFN beta-1a, and glatiramer acetate), monthly-pulsed dose glucocorticoids, as well as azathioprine, mycophenolate mofetil and teriflunomide. Therapies referred to as "intermediate- or high-potency therapy" included natalizumab, rituximab, mitoxantrone, cyclophosphamide, fingolimod and dimethyl fumarate. In order to simplify the nomenclature, in the manuscript “intermediate or high- potency therapies” are mentioned as “high-potency therapies”. Basic clinical and demographic descriptors for the different group are shown in eTable 6.

eAppendix 2. Brain MRI Scans

Brain MRI Scans were acquired on the same 3T GE scanner (GE Medical Systems, Milwaukee, WI) with standardized head positioning and pulse sequences that included: high resolution T1-weighted volume (inversion recovery spoiled gradient-echo, repetition time [TR]/echo time [TE]/inversion time [TI] 5 7/2/400 milliseconds, flip angle 5 88, resolution 5 0.94 3 0.94 3 1mm) with and without gadolinium– diethylenetriamine pentaacetic acid (DPTA); and T2-weighted volume (fast-recovery fastspin-echo [FRFSE], TR/TE 5 2,000/ 81 milliseconds, resolution 5 0.47 3 0.47 3 3mm). Proton density–weighted images were acquired from baseline to year 4 (FRFSE, TR/TE 5 2,000/20 milliseconds, resolution 5 0.47 3 0.47 3 3mm), and fluid-attenuated inversion recovery images (fastspin-echo, TR/TE/TI 5 9,000/126/2,200 milliseconds, resolution 5 0.47 3 0.47 3 3mm) were acquired thereafter. The T2- and T1-weighted images were used to determine MS lesion borders using semiautomated lesion segmentation software (Amira [FEI, Hillsboro, OR] and Lesion Segmentation Toolbox [Structural Brain Mapping Group, Jena, Germany]).

Lesion masks for each time point were created. The lesion masks were then used to subtract MRI lesions from the T1-acquired images. The masked T1-weighted images were used to segment gray matter and white matter structures for volumetric analyses (FreeSurfer). The MS lesion masks were also used to determine the T2 lesion volume (the radiologic burden of disease). Gadolinium-DPTA was administered for the T1 plus contrast enhanced scans, and a neuroradiologist determined the number of gadolinium-enhanced lesions and interpreted all MRI scans to insure safety. Brain volume refers to the absolute value of brain volume assessed at each time point; lesion volume refers to the absolute value of lesion volume assessed at each time point; relative brain volume refers to the percentage of brain volume at a specific time point in reference to the brain volume assessed at the baseline MRI.

eAppendix 3. Statistical Analysis

All statistical analyses were computed using code written in R version 3.4.3 (r-project.org). In all analyses, sNfL levels were natural logarithm (ln)-transformed to meet normal distribution.

For cross-sectional association between sNfL levels and both demographic and clinical variables (age, gender, disease duration, EDSS, disease course, presence of relapse in the 60 days prior to sampling, treatment status – untreated, platform therapy, high potency therapy - and number of HLA- DRB1*15:01 alleles) linear regression models were deployed. This analysis was performed at the first and last time point available for each participant. Estimates were back transformed to the original scale and representing multiplicative effects on the geometric mean of sNfL. Since the natural log transformation of sNfL was the primary outcome, the effects were multiplicative after exponential transformation on the model. After exponential transformation, ln(sNfL) ~ beta0 + beta1*x1 + beta2*x2 becomes sNfL ~ exp(beta0 + beta1*x1 + beta2*x2), and therefore sNfL ~ ebeta0 + beta1*x1 + beta2*x2. A linear regression model was also used to estimate the association between sNfL levels and time from the last clinical exacerbation.

Because it was previously described that sNfL levels increase after a relapse and return to baseline levels after approximately 3 months [2, 3], analysis of the association with relapses was limited to those patients experiencing a relapse within 4 months prior to sampling.

To determine the relationship between sNfL and clinical change over time, linear mixed-effects models with a random intercept were used for two distinct clinical scenarios: (1) we determined the pattern of sNfL level change over time comparing active and inactive subjects based on the presence or absence of clinical exacerbation. This analysis was restricted to participants with CIS and RRMS and including only those from baseline to year 5 of the study, which contain the majority of samples available. (2) We assessed the pattern of sNfL change over time relative to disability worsening (progressors vs. non-progressors) based on clinically significant EDSS change from baseline to years 4 to 6 and confirmed at years 9 to 11. For both analysis covariates included age, gender and disease duration.

A naïve Bayes classifier [4] model was used to evaluate baseline sNfL levels as predictors of the clinical outcome (presence of relapse or sustained EDSS worsening) during different follow- up periods. Age, disease duration and gender were included as covariates. Receiver-operating characteristic (ROC) curves were used to visualize the model performance. ROC curve analysis assesses both sensitivity and specificity performance of a biomarker, and were used to visualize the prognostic value of NfL levels in predicting a clinical outcome. The values for this test range from 0 to 1, wher a value of 1 shows perfect performance and where a value of 0.5 is an uninformative classifier. In general, an AUC of 0.5 suggests no discrimination (i.e., ability to diagnose patients with and without the disease or condition based on the test), 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding [5].

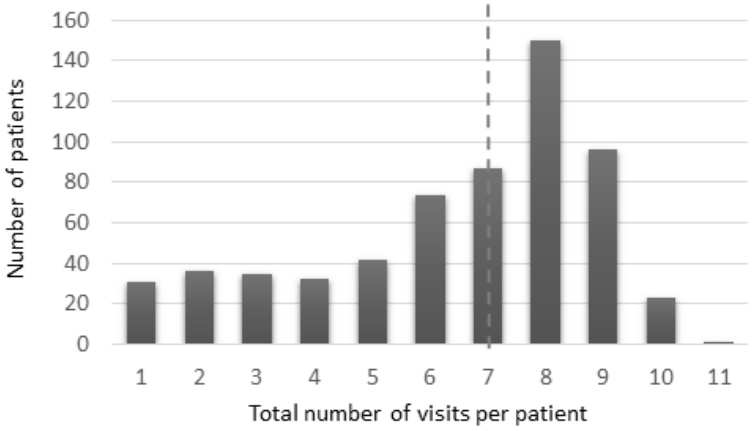
In a separate analysis, samples were categorized into high and low sNfL levels based on different percentiles (80th, 90th, 95th and 99th) generated from all measurements. Using these percentile cutoffs, the association between sNfl levels above versus below various percentiles and clinically meaningful events (presence of relapse or EDSS worsening during the previous year as well as during the subsequent year) was tested using logistic regression models with age, disease duration and gender as covariates. To assess a more homogeneous population, in these analyses only participants with CIS/RRMS were analyzed (total number of samples=2830).

To assess the effect of treatment on sNfL levels, treatment status was stratified as follows: “untreated” included only participants who were not treated over the follow-up period of 3 or 5 years; “platform therapy” included participants who were treated only with platform therapies during the follow-up periods; “intermediate- or high-potency” included participants treated with high-potency therapies or who switched to a high-potency therapy during the designated follow-up periods. The median time for switch to high-potency treatment was 25 months (IQR = 13-37) over the three-year follow-up study and 34 months (IQR = 17-45) over the five-year follow-up study. For this analysis, a linear mixed effect model was applied.

To determine the association between sNfL and MRI markers, generalized estimating equation (GEE) models were used, with the year of follow-up, gender, age, disease duration and the

presence or absence of HLA-DRB1*15:01 allele as covariates. A least square regression model with gender, age at baseline and disease duration at baseline was used to assess baseline sNfL levels as predictors of the percent change in brain fraction from baseline (brain fraction atrophy) at different time points. Finally, to determine the relationship between sNfL and brain fraction over time, GEE models were used. For that purpose we determined the pattern of brain fraction measures over time comparing subjects above and below the established sNfL percentiles at baseline. This analysis was restricted to participants with CIS and RRMS. For these analysis covariates included age, gender and disease duration.

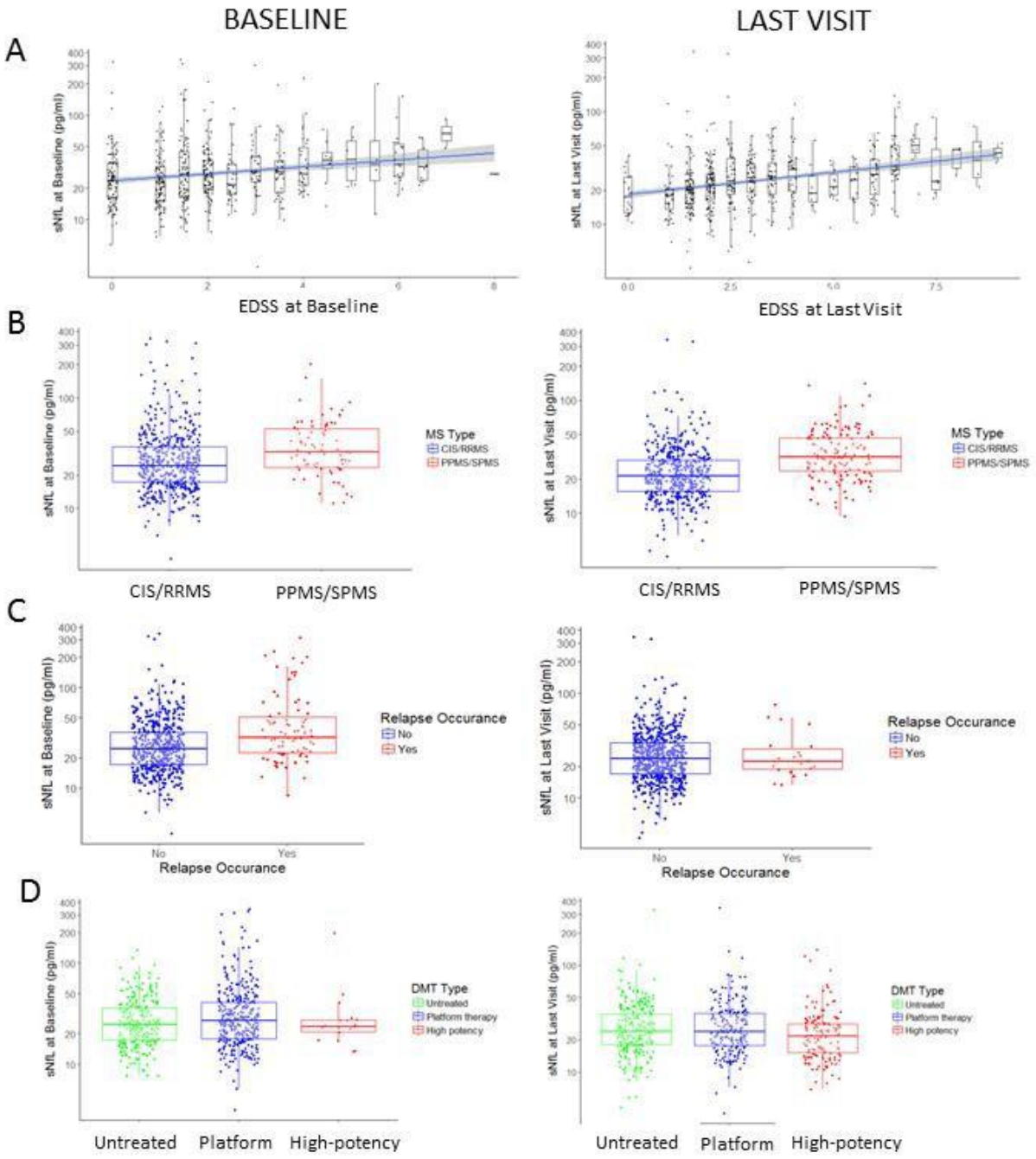
eFigure 1. Visit counts and sample availability



	Visit counts for the longitudinal study													
	BL	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	N
Samples	587	553	498	435	418	435	23	45	299	134	162	264	51	3904
MRI scans	372	367	355	334	334	349	52	30	251	143	166	292	285	3330

Participants were clinically assessed every year during the first 5 years of the study and thereafter re- evaluated at different time points up to 12 years, with a median time of follow-up of 10 years (IQR = 7-11). The histogram shows the frequency of the number of visits per patient with sample available for the study. The dashed line corresponds to the median of samples available per patient. The table below summarizes the total number of samples and MRI scans available for each time point of the study. From the 607 participants, baseline serum samples were available for 587 (97%). Two or more annual samples were available for 576 (95%), and more than half of the participants had at least seven samples available for sNfL measurement (eFigure 1). BL: baseline, Y: year of the study, N: total number of samples.

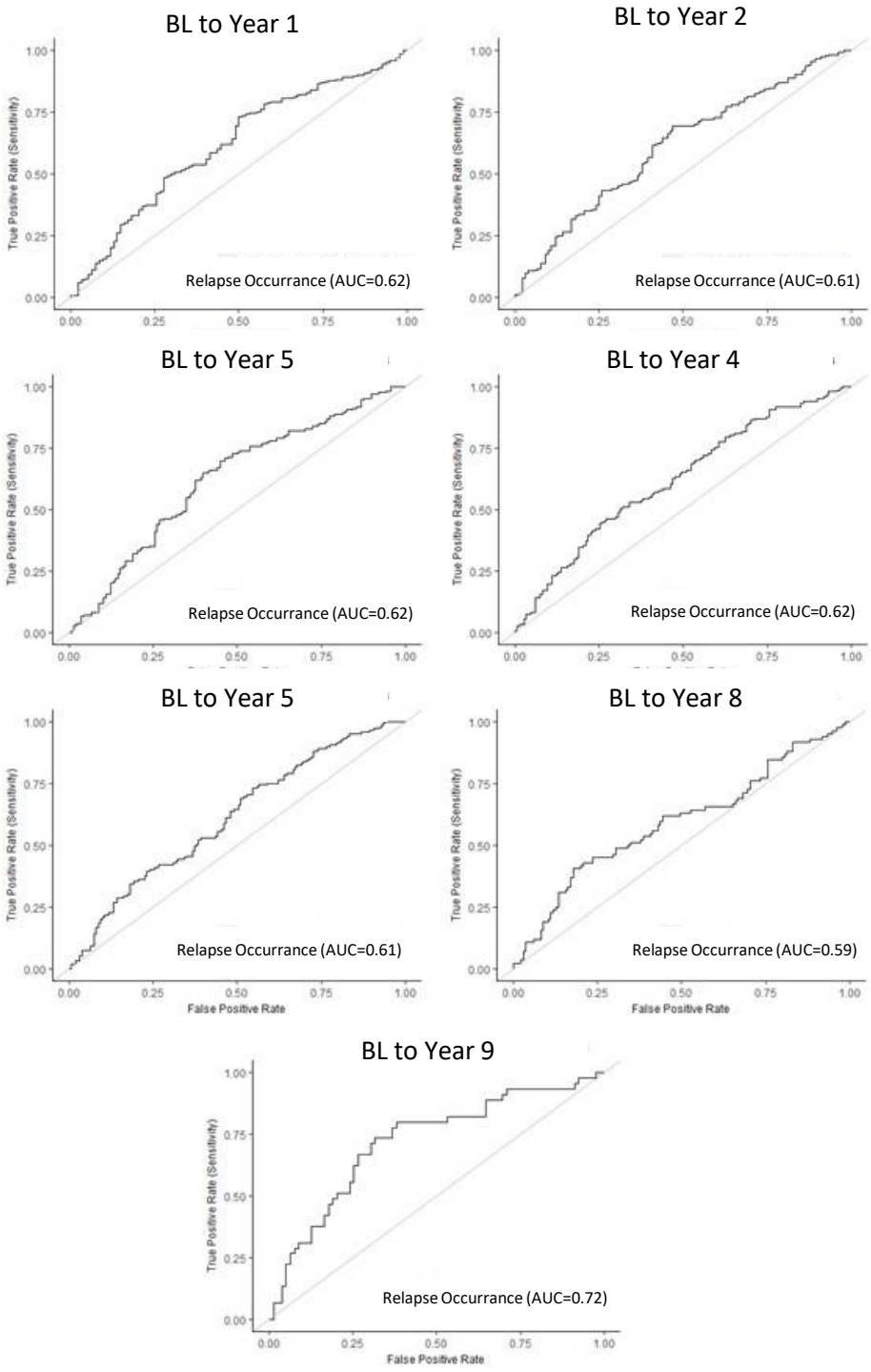
eFigure 2. Associations between NfL levels and clinical variables



(A) Left scatter plot represents the association between sNfL and EDSS at baseline (univariate analysis: p <0.001, multivariate analysis: p = 0.009, see eTable 1). Right scatter plot represents the association between sNfL and EDSS at the last visit available for each patient (univariate analysis: p < 0.001, multivariate analysis: p = 0.006, see eTable 2). Gray band represents the 95% confidence interval. (B) Left boxplot shows median levels of sNfL at baseline in patients with relapsing MS (CIS/RRMS, n=511) compared to patients with progressive forms (PPMS/SPMS, n=75) (univariate analysis: p < 0.001, multivariate analysis: p = 0.228). Right boxplot shows median levels of sNfL at the last visit in patients with relapsing MS (CIS/RRMS, n=455) compared to patients with progressive forms (PPMS/SPMS, n=140) (univariate analysis: p < 0.001, multivariate analysis: p = 0.022). (C) Left boxplot shows median levels of sNfL at baseline in patients not

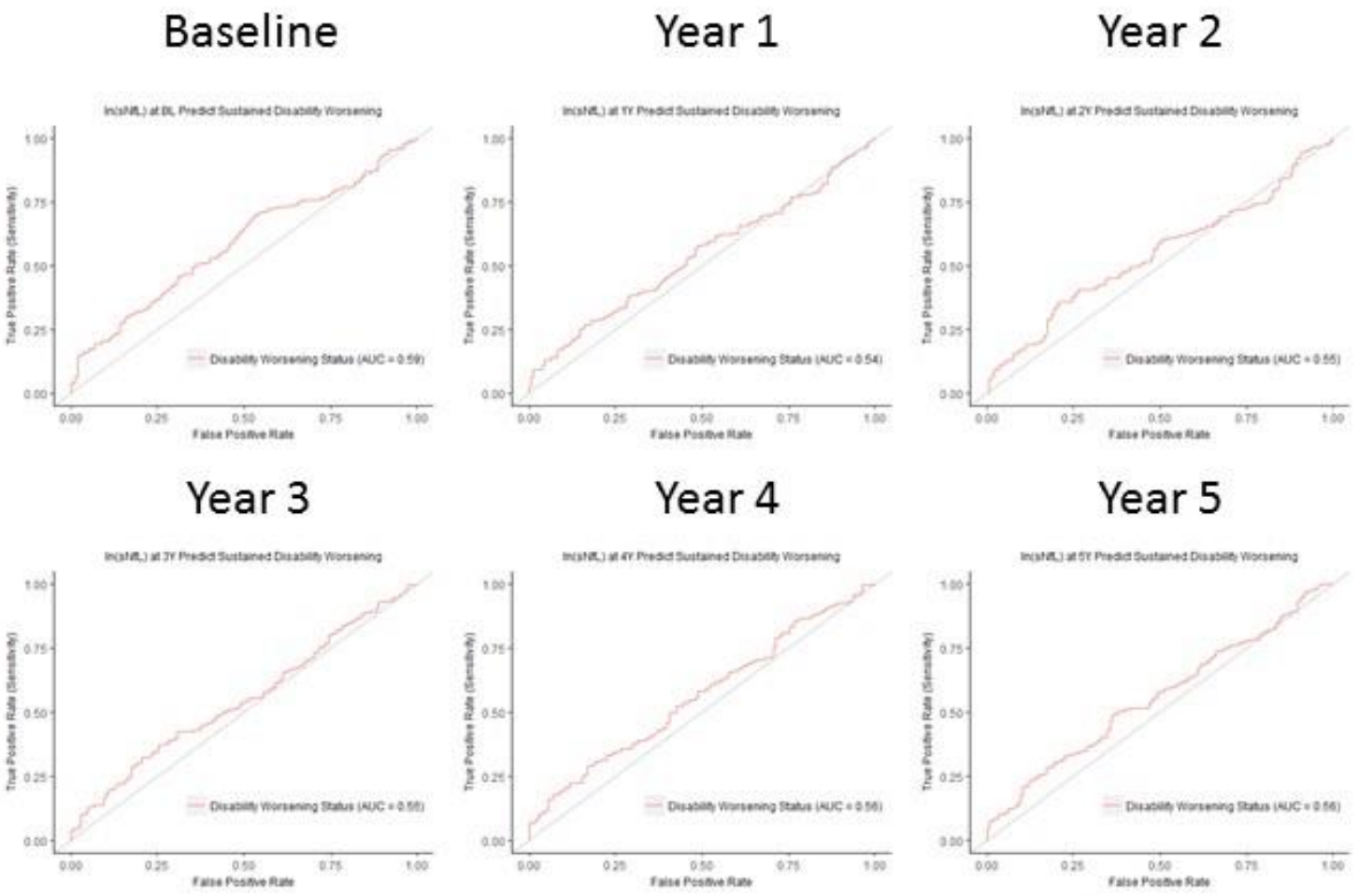
experiencing a relapse in the 90 days prior to sample (n=505) compared to patients experiencing a relapse (n=82) (univariate analysis: $p < 0.001$, multivariate analysis: $p < 0.001$). Right boxplot shows median levels of sNfL at the last visit in patients not experiencing a relapse in the 90 days prior to sample (n=585) compared to patients experiencing a relapse (n=22) (univariate analysis: $p = 0.799$, multivariate analysis: $p = 0.215$). (D) Left boxplot shows median levels of sNfL at baseline in patients untreated (n=227) compared to patients treated with platform therapies (n=342) or treated with high potency therapies (n=18) (univariate analysis: puncorrected = 0.037, pBonferroni = 0.074 and p uncorrected = 0.853, p Bonferroni = 1.000 respectively; multivariate analysis: p uncorrected = 0.027, p Bonferroni = 0.054 and p uncorrected = 0.844, p Bonferroni = 1.000 respectively). Right boxplot shows median levels of sNfL at the last visit in patients untreated (n=247) compared to patients treated with platform therapies (n=209) or treated with high potency therapies (n=151) (univariate analysis: p uncorrected = 0.713, p Bonferroni = 1.000 and p uncorrected = 0.013, p Bonferroni = 0.026 respectively; multivariate analysis: puncorrected = 0.086, pBonferroni = 0.172 and p uncorrected = 0.009, p Bonferroni = 0.018 respectively).

eFigure 3. NfL levels and prediction of relapse



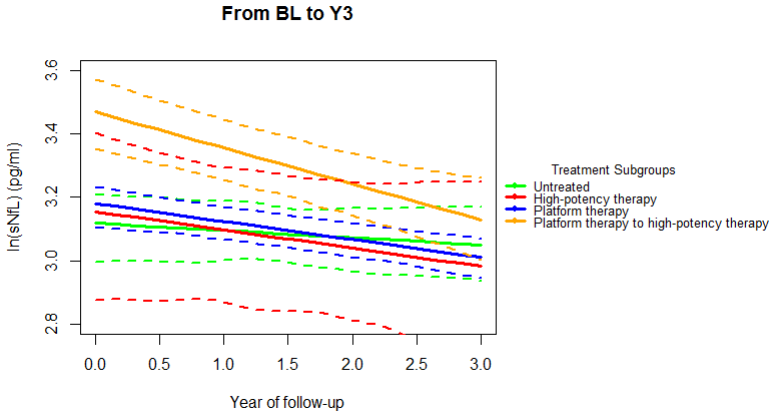
Receiver operating characteristics curves for the prediction of relapse occurrance during different periods based on sNfL levels at baseline of the study. ROC curves show that sNfL levels at baseline are not good predictors of the future disease activity (AUC range = 0.51 – 0.72).

eFigure 4. NfL levels and prediction of disability worsening

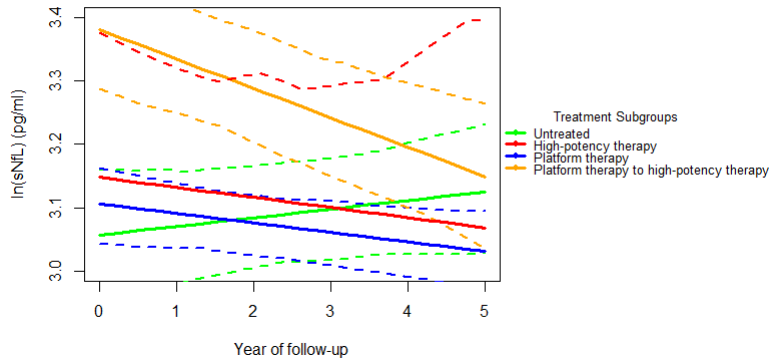


Receiver operating characteristics curves for the prediction of sustained EDSS worsening based on sNfL at different time points. ROC curves show that sNfL levels are not good predictors of disability progression (AUC range = 0.54 – 0.59).

eFigure 5. Treatment effect on NfL levels



	Untreated	Platform Therapy	Platform Therapy to High-potency	High Potency
Untreated (103)	1	-	-	-
Platform Therapy (320)	0.168	1	-	-
Platform Therapy to High-potency (81)	**0.005**	**0.035**	1	-
High Potency (19)	0.561	0.989	0.359	1



	Untreated	Platform Therapy	Platform Therapy to High-potency	High Potency
Untreated (93)	1	-	-	-
Platform Therapy (283)	**0.041**	1	-	-
Platform Therapy to High-potency (94)	**0.0004**	**0.023**	1	-
High Potency (15)	0.418	0.980	0.401	1

Change of sNfL levels over time in participants under different treatment groups: untreated, treated with platform therapies, treated with high-potency therapies or participants that switched from platform to high-potency therapies during the follow-up period. The graphs represent the group means of sNfL over time. The median time of switch to high-potency was 25 months (13-37) for the 3- year follow-up study and 34 months (17-45) for the 5-year follow-up study. Numbers in parenthesis indicate the number of participants in each treatment group. Levels of sNfL show a different rate of change over time in those participants that switch to a high-potency therapy during the follow up period compared to untreated and those treated with platform therapies. Participants treated with platform therapies for 5 years also showed a different rate of sNfL change over time compared to those who were untreated. The analysis includes RRMS, CIS, SPMS and PPMS patients.

eTable 1. Univariate and multivariate models to test the association between sNfL levels at baseline and clinical and demographical variables

Variable (sample Number)	sNfL, Median (IQR), pg/ml	Univariate			Multivariate		
		β	95% CI	p-value	β	95% CI	p-value
Age (587)	-	1.008	1.003 – 1.014	**0.002**	1.007	1.001 – 1.013	**0.031**
Gender							
Female (415)	25.30 (17.70 – 38.50)	-	-	-	-	-	-
Male (172)	25.80 (18.23 – 38.27)	0.984	0.879 – 1.102	0.783	0.973	0.868 – 1.090	0.630
Disease Duration	-	1.005	0.999 – 1.011	0.093	0.996	0.988 – 1.003	0.229
EDSS	-	1.080	1.047 – 1.114	**<0.001**	1.058	1.014 – 1.104	**0.009**
Disease Subtype							
CIS/RRMS (511)	24.40 (17.40 – 36.35)	-	-	-	-	-	-
PPMS/SPMS (75)	32.40 (23.50 – 52.80)	1.312	1.126 – 1.528	**<0.001**	1.128	0.927 – 1.373	0.228
Presence of relapse <90 days prior to sample							
No (505)	24.60 (17.10 – 35.90)	-	-	-	-	-	-
Yes (82)	31.85 (22.43 – 50.60)	1.478	1.279 – 1.707	**<0.001**	1.469	1.266 – 1.704	**<0.001**
DMT							
Untreated (227)	24.50 (17.35 – 36.30)	-	-	-	-	-	-
Platform Therapy (342)	26.95 (17.85 – 41.05)	1.120	1.007 – 1.245	**0.037**	1.129	1.014 – 1.257	**0.027**
High-potency (18)	23.60 (21.02 – 27.32)	1.029	0.760 – 1.394	0.853	0.962	0.656 – 1.411	0.844
HLA DRB1*15:01							
0 Copy (304)	25.80 (16.57 – 40.12)	-	-	-	-	-	-
1 Copies (226)	25.95 (19.30 – 38.90)	1.063	0.952 – 1.188	0.274	1.069	0.960 – 1.190	0.225
2 Copies (34)	23.85 (17.38 – 28.48)	0.905	0.721 – 1.135	0.387	0.872	0.696 – 1.092	0.232

Analysis of the effect of clinical and demographical characteristics on sNfL levels. The number of samples available for each category is indicated in parenthesis. CI = confidence interval. CIS = clinically isolated syndrome, RRMS = relapsing remitting multiple sclerosis, PPMS = primary progressive multiple sclerosis, SPMS = secondary progressive multiple sclerosis, DMT = disease modifying treatment, Platform therapy (interferon (IFN) beta1b, IFN beta1a, and glatiramer acetate, monthly pulsed dose glucocorticoids, azathioprine, mycophenolate mofetil and teriflunomide), High-potency (natalizumab, rituximab, mitoxantrone, cyclophosphamide, fingolimod and dimethyl fumarate), HLA = Human Leukocyte Antigen.

eTable 2. Univariate and multivariate models to test the association between sNfL levels at the last visit available for the study and clinical and demographical variables

Variable (sample Number)	sNfL, Median (IQR), pg/ml	Univariate			Multivariate		
		β	95% CI	p-value	β	95% CI	p-value
Age (607)	-	1.022	1.018 – 1.026	<**0.001**	1.021	1.016 – 1.025	<**0.001**
Gender							
Female (423)	23.70 (17.05 – 33.05)	-	-	-	-	-	-
Male (184)	23.35 (16.55 – 33.55)	0.996	0.907 – 1.094	0.933	0.975	0.896 – 1.062	0.561
Disease Duration	-	1.013	1.008 – 1.017	<**0.001**	0.997	0.992 – 1.002	0.208
EDSS	-	1.095	1.071 – 1.120	<**0.001**	1.045	1.013 – 1.077	**0.006**
Disease Subtype							
CIS/RRMS (455)	21.40 (15.60 – 29.60)	-	-	-	-	-	-
PPMS/SPMS (140)	31.35 (23.55 – 46.35)	1.469	1.331 – 1.621	<**0.001**	1.173	1.023 – 1.344	**0.022**
Presence of relapse <60 days prior to sample							
No (585)	23.60 (16.80 – 33.40)	-	-	-	-	-	-
Yes (22)	22.35 (18.70 – 29.68)	1.031	0.817 – 1.300	0.799	1.139	0.927 – 1.398	0.215
DMT							
Untreated (247)	24.00 (18.10 – 35.05)	-	-	-	-	-	-
Platform Therapy (209)	23.90 (17.80 – 35.40)	1.019	0.922 – 1.125	0.713	1.082	0.989 – 1.185	0.086
High-potency (151)	21.60 (15.05 – 28.40)	0.871	0.780 – 0.971	**0.013**	0.873	0.789 – 0.967	**0.009**
HLA DRB1*15:01							
0 Copy (318)	23.80 (16.25 – 32.88)	-	-	-	-	-	-
1 Copies (229)	23.60 (17.50 – 33.90)	1.025	0.934 – 1.125	0.602	1.046	0.964 – 1.135	0.278
2 Copies (36)	20.60 (16.45 – 29.00)	0.970	0.803 – 1.171	0.750	0.998	0.846 – 1.177	0.982

Analysis of the effect of clinical and demographical characteristics on the dependent variable sNfL levels, which was assessed by univariate linear regression and multivariate linear regression. The number of samples available for each category is indicated in parenthesis. CI = confidence interval. CIS = clinically isolated syndrome, RRMS = relapsing remitting multiple sclerosis, PPMS = primary progressive multiple sclerosis, SPMS = secondary progressive multiple sclerosis, DMT = disease modifying treatment, Platform therapy (interferon (IFN) beta-1b, IFN beta1a, and glatiramer acetate, monthly pulsed dose glucocorticoids, azathioprine, mycophenolate mofetil and teriflunomide), High-potency (natalizumab, rituximab, mitoxantrone, cyclophosphamide, fingolimod and dimethyl fumarate), HLA = Human Leukocyte Antigen. Note that number of subjects in each category differs from the previous table as noted in Table 1.

eTable 3. Association of sNfL levels and past relapse activity and disability worsening

A

Percentile	Samples sNfL above percentile N (%)	Relapses 60 days before sampling			
		OR	95% CI	SE	p-value
70th	854 (30.2)	2.27	1.63 – 3.16	0.38	< 0.001
80th	569 (20.1)	2.64	1.86 – 3.71	0.46	< 0.001
90th	281 (9.9)	3.72	2.52 – 5.40	0.72	< 0.001
95th	142 (5.0)	3.58	2.18 – 5.68	0.87	< 0.001
99th	29 (1.0)	3.78	1.36 – 9.01	1.79	0.005

B

Percentile	Samples sNfL above percentile N (%)	Relapses 1 year before sampling			
		OR	95% CI	SE	p-value
70th	854 (30.2)	2.28	1.86 – 2.79	0.24	< 0.001
80th	569 (20.1)	2.48	1.98 – 3.09	0.28	< 0.001
90th	281 (9.9)	3.19	2.43 – 4.19	0.44	< 0.001
95th	142 (5.0)	3.49	2.43 – 4.99	0.64	< 0.001
99th	29 (1.0)	3.57	1.67 – 7.79	1.39	0.001

C

Percentile	Samples sNfL above percentile N (%)	EDSS worsening in the last year			
		OR	95% CI	SE	p-value
70th	301 (26.3)	1.09	0.79 – 1.49	0.18	0.594
80th	188 (16.4)	0.98	0.67 – 1.42	0.19	0.914
90th	94 (8.2)	1.01	0.60 – 1.64	0.26	0.979
95th	44 (3.8)	1.41	0.70 – 2.69	0.48	0.310
99th	5 (0.4)	0.79	0.04 – 5.38	0.89	0.833

OR= odds ratios, CI= confidence interval, SE= standard error.

Logistic regression models with age, disease duration and gender as covariates were used to assess the association between sNfl levels above versus below various percentiles and clinically meaningful events (relapse or EDSS worsening). Only CIS/RRMS patients were included in the analysis (n=441).

eTable 4. Association of sNfL levels and future relapse activity and disability worsening

A

Percentile	Samples sNfL above percentile N (%)	Relapses 60 days after sampling			
		OR	95% CI	SE	p-value
70th	854 (30.2)	2.09	1.25 – 3.46	0.54	0.004
80th	569 (20.1)	1.98	1.12 – 3.37	0.55	0.015
90th	281 (9.9)	1.90	0.93 – 3.56	0.65	0.058
95th	142 (5.0)	1.11	0.33 – 2.75	0.58	0.850
99th	29 (1.0)	1.09	0.06 – 5.39	1.13	0.931

B

Percentile	Samples sNfL above percentile N (%)	Relapses 1 year after sampling			
		OR	95% CI	SE	p-value
70th	854 (30.2)	1.83	1.43 – 2.33	0.23	<0.001
80th	569 (20.1)	1.67	1.27 – 2.18	0.23	< 0.001
90th	281 (9.9)	1.35	0.94 – 1.90	0.24	0.092
95th	142 (5.0)	1.44	0.90 – 2.22	0.33	0.115
99th	29 (1.0)	1.67	0.64 – 3.85	0.75	0.252

C

Percentile	Samples sNfL above percentile N (%)	EDSS worsening in the next year			
		OR	95% CI	SE	p-value
70th	367 (32.1)	1.22	0.91 – 1.64	0.18	0.184
80th	260 (22.7)	1.23	0.89 – 1.70	0.20	0.198
90th	144 (12.6)	0.99	0.64 – 1.48	0.21	0.944
95th	79 (6.9)	0.63	0.33 – 1.13	0.20	0.141
99th	16 (1.4)	1.03	0.28 – 3.01	0.60	0.960

Logistic regression models with age, disease duration and gender as covariates were used to assess the association between sNfl levels above versus below various percentiles and clinically meaningful events (relapse or EDSS worsening). Only CIS/RRMS patients were included in the analysis (n=441). OR= odds ratios, CI= confidence interval, SE= standard error.

eTable 5. Sensitivity analysis of the association of NfL levels with past and future relapse activity

Percentile	Samples sNfL above percentile N (%)	Relapses 60 days before sampling			
		OR	95% CI	SE	p-value
70th	729 (30.3)	1.79	1.19 – 2.68	0.37	0.005
80th	487 (20.2)	2.29	1.48 – 3.47	0.50	<0.001
90th	242 (10.1)	3.07	1.88 – 4.86	0.74	<0.001
95th	121 (5.0)	5.02	2.89 – 8.38	1.36	<0.001
99th	24 (1.0)	4.21	1.20 – 11.45	2.36	0.010

Percentile	Samples sNfL above percentile N (%)	Relapses 1 year before sampling			
		OR	95% CI	SE	p-value
70th	729 (30.3)	2.00	1.56 – 2.55	0.25	<0.001
80th	487 (20.2)	2.10	1.60 – 2.74	0.29	<0.001
90th	242 (10.1)	2.22	1.59 – 3.06	0.37	<0.001
95th	121 (5.0)	2.82	1.85 – 4.23	0.59	<0.001
99th	24 (1.0)	2.70	1.07 – 6.28	1.20	0.026

Percentile	Samples sNfL above percentile N (%)	Relapses 60 days after sampling			
		OR	95% CI	SE	p-value
70th	729 (30.3)	1.96	1.04 – 3.60	0.62	0.033
80th	487 (20.2)	2.53	1.28 – 4.77	0.84	0.005
90th	242 (10.1)	2.75	1.22 – 5.59	1.05	0.008
95th	121 (5.0)	3.39	1.26 – 7.71	1.54	0.007
99th	24 (1.0)	2.31	0.13 – 11.54	2.39	0.421

Percentile	Samples sNfL above percentile N (%)	Relapses 1 year after sampling			
		OR	95% CI	SE	p-value
70th	729 (30.3)	1.63	1.22 – 2.17	0.24	<0.001
80th	487 (20.2)	1.51	1.08 – 2.08	0.25	0.014
90th	242 (10.1)	1.01	0.63 – 1.57	0.24	0.954
95th	121 (5.0)	1.00	0.51 – 1.78	0.32	1.00
99th	24 (1.0)	0.36	0.02 – 1.72	0.37	0.316

Sensitivity analysis excluding baseline samples. Logistic regression models with age, disease duration and gender as covariates were used to assess the association between sNfl levels above versus below various percentiles and clinically meaningful events (relapse or EDSS worsening). Only CIS/RRMS patients were included in the analysis (n=441). OR= odds ratios, CI= confidence interval, SE= standard error.

eTable 6. Effect of treatment on sNfL levels. Clinical and demographical characteristics of the treatment groups

Analysis from Baseline to Year 3					
Variable	Level	Untreated (n=97)	Platform therapy (n=310)	Intermediate or High potency therapy (n=98)	p-value
Age at exam (years)	mean (SD)	45.1 (10.2)	42.0 (9.6)	41.8 (9.9)	0.016
Gender	F	70 (72.2)	217 (70.0)	58 (59.2)	
	M	27 (27.8)	93 (30.0)	40 (40.8)	0.089
Disease Course	CIS	34 (35.1)	34 (11.0)	3 (3.1)	
	PPMS	4 (4.1)	5 (1.6)	7 (7.1)	
	RRMS	49 (50.5)	254 (81.9)	69 (70.4)	
	SPMS	10 (10.3)	17 (5.5)	19 (19.4)	<0.001
Disease Duration (years)	mean (SD)	9.2 (10.8)	7.8 (7.9)	10.6 (8.0)	0.013
EDSS	mean (SD)	1.7 (1.5)	1.7 (1.5)	3.0 (1.7)	<0.001

Analysis from Baseline to Year 5					
Variable	Level	Untreated (n=97)	Platform therapy (n=310)	Intermediate or High potency therapy (n=98)	p-value
Age at exam (years)	mean (SD)	45.4 (10.4)	41.6 (9.4)	41.8 (9.2)	0.005
Gender	F	65 (73.0)	192 (70.3)	67 (62.0)	
	M	24 (27.0)	81 (29.7)	41 (38.0)	0.188
Disease Course	CIS	32 (36.0)	34 (12.5)	3 (2.8)	
	PPMS	4 (4.5)	5 (1.8)	5 (4.6)	
	RRMS	44 (49.4)	222 (81.3)	82 (75.9)	
	SPMS	9 (10.1)	12 (4.4)	18 (16.7)	<0.001
Disease Duration (years)	mean (SD)	9.2 (10.7)	7.6 (8.0)	9.8 (8.2)	0.056
EDSS	mean (SD)	1.7 (1.5)	1.6 (1.5)	2.6 (1.7)	<0.001

Values shown as mean (standard deviation) or count (percentage). CIS = clinically isolated syndrome. PPMS = primary progressive multiple sclerosis, RRMS = relapsing remitting multiple sclerosis, SPMS = secondary progressive multiple sclerosis, Platform treatment (interferon (IFN) beta1b, IFN beta1a, and glatiramer acetate, monthly pulsed dose glucocorticoids, azathioprine, mycophenolate mofetil and teriflunomide), Intermediate- or high-potency treatment (natalizumab, rituximab, mitoxantrone, cyclophosphamide, fingolimod and dimethyl fumarate).

eTable 7. Association of NfL levels and MRI markers

Brain Volume			
	Estimate	SE	p-value
Intercept	1675.59	16.39	<0.001
Follow-up year	-9.14	0.31	<0.001
Ln(sNfL)	1.60	1.33	0.228
Age	-3.28	0.37	<0.001
Disease duration	-1.91	0.40	<0.001
Gender (male)	-25.54	6.60	<0.001
HLA-DRB1*15:01	-8.42	6.10	0.167
Lesion Volume			
Intercept	11.28	6.55	0.085
Follow-up year	-0.34	0.21	0.116
Ln(sNfL)	5.64	1.70	<0.001
Age at exam	-0.35	0.13	<0.001
Disease duration	0.71	0.18	<0.001
Gender (male)	-1.41	2.02	0.486
HLA-DRB1*15:01	-1.07	1.83	0.560
Brain Fraction			
Intercept	0.9980	0.0018	<0.001
Follow-up year	-0.0008	0.0000	<0.001
Ln(sNfL)	0.0003	0.0001	<0.001
Age at exam	-0.0003	0.0000	<0.001
Disease duration	-0.0003	0.0001	<0.001
Gender (male)	-0.0040	0.0010	<0.001
HLA-DRB1*15:01	-0.0016	0.0008	0.063

Analysis of the effect of clinical and demographical characteristics on MRI outcomes. Generalized estimating equation (GEE) models were used, with the year of follow-up, Ln(sNfl) at each time point, gender, age at baseline, disease duration at baseline and the presence or absence of HLA-DRB1*15:01 allele as covariates. SE = standard error, HLA = Human Leukocyte Antigen.

eTable 8. Baseline NfL levels as predictors of brain fraction atrophy

	BFA at year 1			BFA at year 2			BFA at year 3			BFA at year 4		
	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept	-0.006	0.044	0.886	0.036	0.064	0.572	0.109	0.098	0.265	0.222	0.129	0.085
Ln(sNfL)	-0.040	0.011	**<0.001**	-0.096	0.015	**<0.001**	-0.110	0.023	**<0.001**	-0.164	0.030	**<0.001**
Age	0.001	0.001	0.178	0.002	0.001	**0.040**	0.001	0.002	0.411	0.001	0.002	0.558
Disease duration	0.0002	0.001	0.820	-0.001	0.001	0.279	-0.005	0.002	**0.011**	-0.006	0.003	**0.023**
Gender (male)	-0.017	0.014	0.234	-0.038	0.021	0.066	-0.064	0.032	**0.044**	-0.124	0.040	**0.002**

	BFA at year 5			BFA at year 8			BFA at year 9		
	β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept	0.267	0.132	**0.045**	0.408	0.220	0.065	0.294	0.328	0.371
Ln(sNfL)	-0.198	0.031	**<0.001**	-0.286	0.054	**<0.001**	-0.319	0.085	**<0.001**
Age	0.002	0.002	0.536	-0.001	0.004	0.814	0.002	0.006	0.761
Disease duration	-0.011	0.003	**<0.001**	-0.015	0.004	**<0.001**	-0.017	0.008	**0.039**
Gender (male)	-0.121	0.042	**0.005**	-0.090	0.070	0.198	-0.157	0.116	0.178

Analysis of the effect of clinical and demographical characteristics on brain atrophy at different follow-up times A least square regression model with gender, age at baseline and disease duration at baseline was used to assess baseline sNfL levels as predictors of brain fraction atrophy (BFA). SE = standard error.

eTable 9. Baseline NfL percentiles as predictors of brain fraction atrophy

	Brain Fraction					
	80th percentile			90th percentile		
	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	0.99700	0.00195	< 2e-16 ***	0.99600	0.00193	< 2e-16 ***
Percentile (below)	0.00009	0.00079	0.90465	0.00132	0.00100	0.18609
Years	-0.00099	0.00007	< 2e-16 ***	-0.00108	0.00009	< 2e-16 ***
Age (at baseline)	-0.00025	0.00005	5.6e-08 ***	-0.00026	0.00005	2.0e-08 ***
Disease duration (at baseline)	-0.00036	0.00006	1.5e-08 ***	-0.00037	0.00006	5.5e-09 ***
Gender (male)	-0.00405	0.00103	8.8e-05 ***	-0.00392	0.00103	0.00014 ***
Percentile (below)*Years	0.00028	0.00007	0.00015 ***	0.00035	0.00010	0.00027 ***

	Brain Fraction					
	95th percentile			99th percentile		
	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	0.99500	0.00202	< 2e-16 ***	0.99600	0.00251	< 2e-16 ***
Percentile (below)	0.00304	0.00133	0.0218 *	0.00147	0.00212	0.49
Years	-0.00110	0.00011	< 2e-16 ***	-0.00094	0.00011	< 2e-16 ***
Age (at baseline)	-0.00027	0.00005	6.3e-09 ***	-0.00027	0.00005	1.8e-08 ***
Disease duration (at baseline)	-0.00037	0.00006	3.2e-09 ***	-0.00036	0.00006	1.3e-08 ***
Gender (male)	-0.00410	0.00101	5.1e-05 ***	-0.00411	0.00104	7.7e-05 ***
Percentile (below)*Years	0.00035	0.00011	0.0023 **	0.00015	0.00012	0.2

Analysis of the effect of different percentile sNfL levels at baseline and brain fraction atrophy over time. Generalized estimating equation (GEE) models were used, with the gender, baseline age and baseline disease duration as covariates. Years represent the effect of follow-up time. Percentile*Years represents the interaction between percentiles group and year of follow up. SE = standard error.

eTable 10. Baseline NfL levels as predictors of brain atrophy variance

BFA Variance	1Y	2Y	3Y	4Y	5Y	8Y	9Y	10Y
Brain fraction atrophy ln(sNfL) at BL	3.5%	7.8%	6.5%	8.1%	11.0%	10.6%	11.0%	11.6%
Brain fraction atrophy. Full model	4.5%	10.1%	9.7%	12.6%	17.9%	17.0%	15.6%	18.3%

The association of baseline NfL levels with brain fraction atrophy (percent change in brain fraction relative to baseline) was assessed by linear regression model: Brain fraction atrophy (BFA) ~ ln(sNfL) at baseline + sex + disease duration at baseline. The variances in BFA predicted from the variable ln(sNfL) and the full model were determined by Analysis of variance (ANOVA) and R-squared respectively. Age at baseline was not a significant predictor.

eReferences

1. University of California San Francisco MS-EPIT Team, Cree BA, Gourraud PA, Oksenberg JR, Bevan C, Crabtree-Hartman E, Gelfand JM, Goodin DS, Graves J, Green AJ, et al: Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol* 2016, 80:499- 510.
2. Lycke JN, Karlsson JE, Andersen O, Rosengren LE: Neurofilament protein in cerebrospinal fluid: a potential marker of activity in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1998, 64:402-404.
3. Teunissen CE, Iacobaeus E, Khademi M, Brundin L, Norgren N, Koel-Simmelink MJ, Schepens M, Bouwman F, Twaalfhoven HA, Blom HJ, et al: Combination of CSF N-acetylaspartate and neurofilaments in multiple sclerosis. *Neurology* 2009, 72:1322-1329.
4. Webb G.I. (2011) Naïve Bayes. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA
5. Hosmer DWL, S.: Applied Logistic Regression. In *Applied Logistic Regression. Volume,* 2nd edition. Edited by John Wiley and Sons NY, NY (2000)