ARTICLE OPEN ACCESS

Blood neurofilament light levels segregate treatment effects in multiple sclerosis

Bénédicte Delcoigne, PhD, Ali Manouchehrinia, PhD, Christian Barro, MD, Pascal Benkert, PhD,
Zuzanna Michalak, PhD, Ludwig Kappos, MD, David Leppert, MD, Jon A. Tsai, MD, Tatiana Plavina, PhD,
Bernd C. Kieseier, MD, Jan Lycke, MD, Lars Alfredsson, PhD, Ingrid Kockum, PhD, Jens Kuhle, MD,*
Tomas Olsson, MD,* and Fredrik Piehl, MD*

Correspondence
Dr. Delcoigne
benedicte.delcoigne@ki.se

Abstract

Objective

To determine factors (including the role of specific disease modulatory treatments [DMTs]) associated with (1) baseline, (2) on-treatment, and (3) change (from treatment start to on-treatment assessment) in plasma neurofilament light chain (pNfL) concentrations in relapsing-remitting multiple sclerosis (RRMS).

Methods

Data including blood samples analyses and long-term clinical follow-up information for 1,261 Swedish patients with RRMS starting novel DMTs were analyzed using linear regressions to model pNfL and changes in pNfL concentrations as a function of clinical variables and DMTs (alemtuzumab, dimethyl fumarate, fingolimod, natalizumab, rituximab, and teriflunomide).

Results

The baseline pNfL concentration was positively associated with relapse rate, Expanded Disability Status Scale score, Age-Related MS Severity Score, and MS Impact Score (MSIS-29), and negatively associated with Symbol Digit Modalities Test performance and the number of previously used DMTs. All analyses, which used inverse propensity score weighting to correct for differences in baseline factors at DMT start, highlighted that both the reduction in pNfL concentration from baseline to on-treatment measurement and the on-treatment pNfL level differed across DMTs. Patients starting alemtuzumab displayed the highest reduction in pNfL concentration and lowest on-treatment pNfL concentrations, while those starting teriflunomide had the smallest decrease and highest on-treatment levels, but also starting from lower values. Both on-treatment pNfL and decrease in pNfL concentrations were highly dependent on baseline concentrations.

Conclusion

Choice of DMT in RRMS is significantly associated with degree of reduction in pNfL, which supports a role for pNfL as a drug response marker.

Glossary

ALM = alemtuzumab; **ARMSS** = Age-Related MS Severity Score; **CI** = confidence interval; **DMF** = dimethyl fumarate; **DMT** = disease-modifying therapies; **EDSS** = Expanded Disability Status Scale; **FGL** = fingolimod; **GA** = glatiramer acetate; **IFN** = interferon; **IMSE** = Immunomodulation and Multiple Sclerosis Epidemiology; **LLoQ** = lower limit of quantification; **MS** = multiple sclerosis; **MSIS-29** = MS Impact Score; **MSSS** = Multiple Sclerosis Severity Score; **NfL** = neurofilament light; **NTZ** = natalizumab; **pNfL** = neurofilament light in plasma; **PS** = propensity scores; **RRMS** = relapsing-remitting multiple sclerosis; **RTX** = rituximab; **SDMT** = Symbol Digit Modalities Test; **sNfL** = neurofilament light in serum; **TFL** = teriflunomide.

Accumulating evidence supports the notion that permanent loss of neurologic functions in multiple sclerosis (MS) is primarily correlated with the degree of damage to nerve tracts rather than degree of demyelination.[1–3] However, due to the reserve capacity of the CNS, critical levels of nerve damage may take years to appear as clinical disability. The observation that disease-modifying therapies (DMT) used in relapsing-remitting MS (RRMS) differently affect important long-term clinical outcomes underscores a need for more sensitive measures of core disease pathologic mechanisms.[4] MRI is the only acknowledged biomarker for disease progression and different volumetric atrophy measures have been associated with risk of developing increasing disability.[5–7] However, such measures are insensitive to changes over shorter time periods in individual patients. Moreover, spinal cord pathology, a major driver of clinical disability, is not routinely assessed. Among different soluble markers for neuroaxonal damage, neurofilaments have emerged as promising candidates in a range of diseases.[8] Although not specific for disease processes operating solely in MS, the potential value in this condition is especially high since it may be used to monitor treatment effects. Most published studies on neurofilament light (NfL) and effects of DMTs have measured concentrations of NfL in CSF focusing on a single or a few DMTs.[9–12] More recently, improvements in assay sensitivity have made it possible to reliably determine NfL in serum (sNfL) or plasma (pNfL) at concentrations seen in healthy controls. Such studies have reported a correlation between baseline levels of pNfL/sNfL and measures of clinical disease activity including development of sustained disability, brain atrophy, signs of nerve tissue damage, and long-term clinical disability outcomes.[13–15] Treatment effects have been reported by several authors.[14,16] Disanto et al.[14] studied 2 Swiss cohorts of patients with MS in which the effects of a limited number of DMTs on NfL were reported. In this study, the decrease in sNfL after initiation of DMT was of similar magnitude across all DMTs, but confidence intervals (CIs) were large due to the small size of the study population. Similarly, Novakova et al.[16] reported a Swedish MS cohort in which start of DMT resulted in lowered sNfL levels, also correlating with CSF NfL concentrations, across all different DMTs, but with low power to address effect size of specific DMTs. Thus so far there is a relative paucity of well-powered studies specifically addressing treatment effects across multiple DMTs in real-world cohorts of patients. The aim of this study was to address treatment effects across multiple DMTs through the measurement of blood NfL at 2 time points in patients selected within a large cohort of patients with RRMS

initiating DMT in context of a nationwide, population-based follow-up program for all newer MS DMTs.

Methods

Patient selection and sample collection

The Immunomodulation and Multiple Sclerosis Epidemiology study (IMSE) is a comprehensive nationwide Swedish post-approval program of patients starting newer MS DMTs, coupled with sampling of blood at initiation of therapy and at follow-up. Samples were collected from patients included in IMSE as well as in the Epidemiologic Investigation of MS and Stockholm Prospective Assessment of MS. We analyzed data for 1,139 patients with RRMS initiating alemtuzumab (ALM, n = 89), dimethyl fumarate (DMF, n = 339), fingolimod (FGL, n = 275), natalizumab (NTZ, n = 284), or teriflunomide (TFL, n = 152). Inclusion criteria comprised a baseline sample within a month prior to day of initiation of DMT and a subsequent treatment duration of >4 months. Most patients (1,052) provided 2 samples (at treatment start and on treatment [absolute range 4–24 months]). Seventeen patients (4%) contributed samples for more than 1 DMT. A follow-up program similar to IMSE was recently started for rituximab (RTX); however, only 11 of 122 analyzed patients had a sample before starting therapy. The total number of patients included in this study is thus 1,261. A total of 1,026 population-based controls included in the study by Manouchehrinia et al.[15] was used to calculate age-adjusted pNfL reference curves.

NfL analyses

pNfL concentrations were determined using antibodies from UmanDiagnostics (Umeå, Sweden) and the SIMOA Immunoassay using the Quanterix Kit (Quanterix, Lexington, MA). All samples from different DMTs were analyzed with blinding for treatment or clinical information. The lower limit of quantification (LLoQ) was 1.95 pg/mL. All measurements were duplicated and were above the LLoQ, with interassay and intra-assay coefficients of variation of ≤10%.

Clinical variables collection

All IMSE patients attend regular medical visits where clinical assessments are carried out and recorded through the Swedish MS registry. In addition to general demographics (age at DMT start, age at MS onset, and sex), we had access to the dates of relapses (if any) before DMT start, the type of previous DMTs (if any) with start and stop dates and the reason for stopping, as well as clinical assessments: Expanded Disability Status Scale

(EDSS), further transformed into the Age-Related MS Severity Score (ARMSS; an alternative to the Multiple Sclerosis Severity Score [MSSS] based on the patient's age at the time of assessment[17]); the MS Impact Score (MSIS-29), divided into its physical and psychological domains; and the Symbol Digit Modalities Test (SDMT) score.

Standard protocol approvals, registrations, and patient consents

The study was approved by the regional vetting board of Stockholm under permits 2006/845-31/1 2011/641-31/4, 2009/2017-31/2, and 04-252/1-4, with written informed consent from all participants.

Statistical analyses

Variables preparation

For all analyses, we log-transformed pNfL levels to increase the normality of the distribution. We also normalized the log-pNfL values to age 40 (log-pNfL$_{N40}$), by using the linear relationship between increasing log-pNfL and age in a large population-based control sample (i.e., log-pNfL$_{N40}$ = log [pNfL] − 0.02115 [age at DMT start − 40]).[15] This normalization implies that a difference between 2 pNfL$_{N40}$ measures cannot be attributable to a difference in ages. We calculated the number of relapses in the year preceding DMT start and the number of previous DMTs (β-interferons/glatiramer acetate [IFN/GA], ALM, DMF, FGL, NTZ, RTX, and TFL) since disease onset for each patient and these 2 variables were considered as numerical. We also created a 3-category variable denoting treatment status at start of the new DMT by including a washout period (time span between stop date of previous DMT and start of new DMT) of at least 1 month for IFN/GA, DMF, and TFL, at least 3 months for FGL and NTZ, and 6 months or more for RTX (none of the patients had switched from ALM). Patients were dichotomized as being treated with IFN/GA or with one of the other DMTs if washout periods had been shorter.

Baseline log-pNfL levels analysis

We analyzed the log-pNfL levels at baseline (without age normalization) with linear models. Initially, we used univariable linear models with log-pNfL levels as the dependent variables and each of the variables measured at baseline (i.e., DMT start) as the independent variables to explore the correlation among log-pNfL levels, clinical variables, and patient characteristics. In a second step, we used a best subset selection approach to determine which subset of the baseline variables contributed most to explaining the variability of the pNfL levels.[18] The tested variables included the number of previous DMTs, treatment status just before DMT start, sex, age at disease onset, disease duration, age at DMT start, number of relapses during the year before DMT start, EDSS, ARMSS, MSIS-29 (physical and psychological scales), and SDMT, all these being measured at DMT start.

Propensity score estimation

In order to balance the DMT groups, we calculated individual DMT propensity scores (PS), i.e., the probability to be treated with a specific DMT.[19,20] We used a multinomial logistic model with ALM, DMF, FGL, NTZ, and TFL as the dependent variable, while the independent variables included all variables measured at DMT start, including log-pNfL$_{N40}$. Several combinations of these variables were tested including interaction terms or transformed scales of variables. The ability of the inverse of the PS in reducing differences between DMT groups in baseline log-pNfL$_{N40}$ values, assessed by measuring the standardized differences between the mean log-pNfL$_{N40}$ values of each DMT group and the overall mean, depended on the input variables. Among different models tested, the one resulting in the smallest average of the standardized differences was selected.[19–21] In the subsequent analyses, we used weights that were calculated by using the inverse of the PS. However, individual weights were limited to the 0.995th percentile of their distribution in order to prevent disproportionate effects on the analytical model.[20] We excluded RTX from PS analyses since baseline pNfL values were available only for a small minority.

Changes in log-pNfL$_{N40}$ levels analysis

We used a graphical approach to describe changes in log-pNfL$_{N40}$ levels from DMT start to follow up (4–24 months later) using unweighted means of the log-pNfL$_{N40}$ across different DMTs, and subsequently, values weighted by the inverse of the PS. As the main question was to assess if different DMTs were significantly associated with degree of reduction of pNfL$_{N40}$ concentrations, we calculated the delta pNfL$_{N40}$ (i.e., change in log-pNfL$_{N40}$ levels). We used a weighted linear model with delta as the dependent variable and the DMTs as the independent variable, using weights obtained by inversing the PS, and further adjusted for other baseline covariates to remove potential residual confounding.[22] Criteria to retain a variable included percentage of the explained variance, and how much the additional variable modified the estimates for the DMTs. In an additional sensitivity analysis, we stratified on the quintiles of the PS instead of using weights. We also analyzed how the changes in log-pNfL$_{N40}$ correlated with the changes in EDSS, ARMSS, MSIS-29, and SDMT using univariable models.

Additional supporting analyses

As RTX was excluded from the analyses using PS, we also modeled the log-pNfL$_{N40}$ on treatment, without using PS but adjusting the analyses for patient characteristics using linear models. In parallel, the log-pNfL$_{N40}$ on treatment without RTX but using PS was also modeled.

Data availability

Data related to the current article are available from Tomas Olsson, Karolinska Institutet. To share data from the Swedish MS registry, a data transfer agreement needs to be completed between Karolinska Institutet and the institution requesting data access. This is in accordance with data protection legislation in Europe (General Data Protection Regulation). Persons interested in obtaining access to the data should contact Tomas Olsson at tomas.olsson@ki.se.

Table 1 Baseline and on-treatment characteristics of patients in the 6 disease-modifying therapy (DMT) groups

Variables	Median (p25–p75) or n (%)						p Value
	ALM	DMF	FGL	NTZ	RTX	TFL	
No.	89	339	275	284	122	152	
Women,[a] n (%)	56 (63)	251 (74)	187 (68)	204 (72)	84 (69)	112 (74)	0.18
Baseline values							
Age, y[a,b]	33 (28–39)	42 (34–49)	38 (32–44)	37 (30–44)	39 (32–48)	45 (40–50)	≤0.001
Age at MS onset, y[b,c]	26 (23–32)	32 (25–40)	28 (23–35)	28 (23–35)	31.5 (24–37)	34.5 (29–40)	≤0.001
Disease duration, y[b,c]	5 (2–10)	7 (2–14)	8 (4–13)	7 (3–13)	3 (3–8)	10 (4–15)	0.002
No. of DMTs[a,d]	2 (1–3)	1 (1–1)	1 (1–2)	1 (1–1)	1 (1–1)	1 (1–2)	≤0.001
EDSS[e]	2 (1.5–3)	1.5 (1–2.5)	2 (1–2.5)	2.5 (2–3.5)	2 (1–3)	2 (1–2.5)	≤0.001
ARMSS[e]	4.9 (2.8–6.4)	2.8 (1.0–4.5)	3.9 (1.9–5.8)	5.4 (3.8–7.0)	3.7 (1.9–5.0)	2.7 (1.1–4.2)	≤0.001
MSIS-29 physical[c]	1.6 (1.3–2.6)	1.5 (1.2–2.1)	1.5 (1.2–2.2)	2.1 (1.5–2.5)	1.4 (1.1–2.2)	1.4 (1.1–2.1)	≤0.001
MSIS-29 psychological[c]	2.1 (1.6–3.1)	2.0 (1.3–2.7)	2.0 (1.4–2.9)	2.4 (1.7–3.1)	2.1 (1.6–2.8)	1.7 (1.2–2.4)	≤0.001
SDMT[f]	60.5 (52–69)	51 (45–59)	51 (43–58)	50 (42–57)	53 (49–62)	52 (46–57)	≤0.001
Relapses[f,g] (% yes)	50%	42%	44%	55%	34%[h]	29%	≤0.001
Treatment history							
Previous treatment (independent of washout)							
No previous treatment	13 (15)	69 (20)	21 (8)	28 (10)	25 (37)[i]	28 (18)	≤0.001
1st line (IFN/GA, TFL, DMF)[a]	13 (15)	265 (78)	231 (84)	248 (87)	28 (41)[i]	117 (77)	≤0.001
2nd line (FGL, NTZ, RTX)[a]	63 (71)	5 (1)	23 (8)	8 (3)	15 (22)[i]	7 (5)	≤0.001
NfL values							
NfL_{N40} baseline[a]	10.5 (6.3–24.8)	11.1 (8.2–15.6)	12.3 (8.7–16.9)	15.5 (9.9–26.9)	12.3[j] (9.7–18.2)	9.0 (7.0–12.2)	≤0.001
NfL_{N40} on treatment[c]	6.9 (5.4–8.8)	8.3 (6.8–10.7)	9.6 (7.6–11.8)	8.7 (7.3–11.8)	9.6 (7.9–11.5)	10.0 (7.2–13)	≤0.001
Time DMT start to 2nd NfL assessment[c,d]	369 (357–377)	366 (364–388)	375 (359–395)	370 (353–389)	379 (361–543)	357 (201–381)	≤0.001

Abbreviations: ALM = alemtuzumab; ARMSS = Age-Related MS Severity Score; DMF = dimethyl fumarate; EDSS = Expanded Disability Status Scale; FGL = fingolimod; GA = glatiramer acetate; IFN = interferon; MS = multiple sclerosis; MSIS-29 = MS Impact Score; NfL = neurofilament light; NTZ = natalizumab; RTX = rituximab; SDMT = Symbol Digit Modalities Test; TFL = teriflunomide.
The last column informs whether the patients significantly differ between groups for each variable (1-way analysis of variance for continuous variables and χ^2 for categorical variables).
[a] 0% missing.
[b] In rounded years.
[c] Less than 10% missing.
[d] Previous to baseline.
[e] 26% to 40% missing.
[f] 11% to 25% missing.
[g] During the year before DMT start.
[h] Calculated on 29 values.
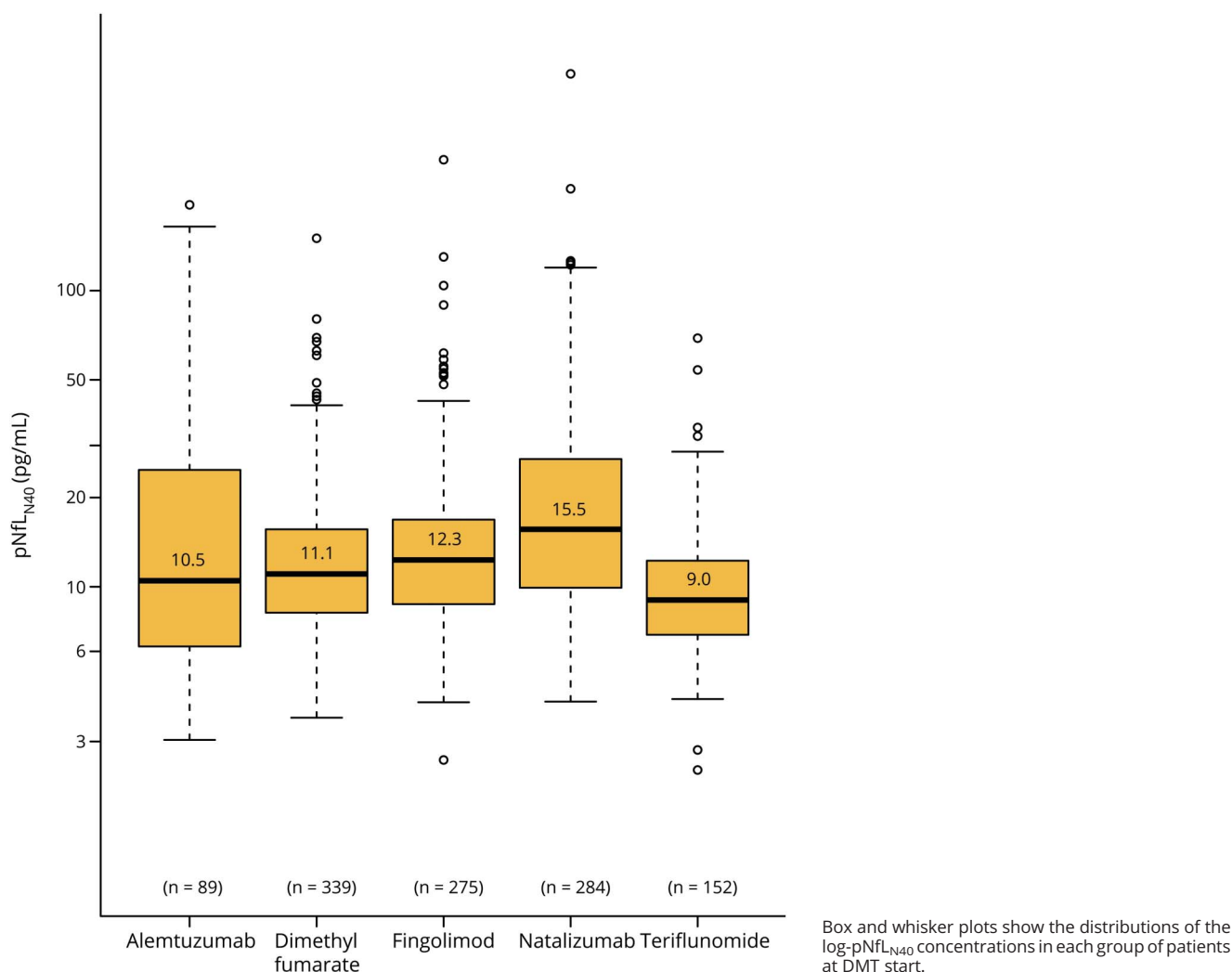[i] Calculated on 68 values.
[j] Calculated on 11 values.

Results

Baseline characteristics

Data on baseline patient characteristics at therapy initiation are presented in table 1. There were large differences between DMT groups, where for example those starting TFL were older both at disease onset and at therapy initiation, had lower MSIS-29 and ARMSS values, and had a longer disease duration compared to other DMT groups (table 1). From a disease severity perspective, NTZ starters were characterized by both higher EDSS and MSIS-29 scores as well as higher relapse activity compared to other groups. These differences

Figure 1 Baseline log–neurofilament light in plasma (pNfL)$_{N40}$ levels in groups starting different disease-modifying therapies (DMTs) (with median and 25th and 75th percentiles)

Box and whisker plots show the distributions of the log-pNfL$_{N40}$ concentrations in each group of patients at DMT start.

were mirrored in both baseline pNfL (data not shown) and baseline pNfL$_{N40}$ concentrations (table 1 and figure 1).

Modeling baseline log pNfL

The pNfL values displayed a skewed distribution and were log transformed. We then modeled log-pNfL levels at baseline (without age normalization) with a linear model. As most of the variables displayed a fluctuating degree of association with the pNfL values and also interacted, we used a best subset selection to model pNfL variability across groups. The back transformed estimates $(\exp[\beta])$ are given in table 2 for both the univariable and multivariable models. The pNfL levels increased with EDSS, ARMSS, MSIS-29 (physical and psychological scales), and number of relapses before DMT start, and decreased with SDMT scores and number of previous DMTs.

Propensity scores

The variables retained for modeling the PS through the multinomial logistic model of the 5 DMTs, excluding RTX,

were selected after testing several combinations of the baseline variables, retaining the model with the smallest average standardized difference. This model included the baseline pNfL$_{N40}$ level, ARMSS, EDSS, SDMT, age at disease onset, the number of previous DMTs, the treatment status just before starting the new DMT, and the number of relapses during the year before DMT start. With these variables, the average of the standardized absolute distances for log-pNfL$_{N40}$ dropped from 0.24 before weighting to 0.05 after weighting (figure 2).

Changes in log-pNfL$_{N40}$ levels analysis

The changes in log-pNfL$_{N40}$ levels between baseline and on treatment are presented in figure 3, both for the unweighted values (figure 3A) and the values weighted with the inverse of the PS (figure 3B). Despite PS weighting, some differences between DMTs remained, suggesting residual effects of factors not accounted for. The estimates from both the unweighted and weighted linear regression models with delta

Table 2 Univariable and multivariable estimates and associated *p* values from a linear model of the baseline log–neurofilament light in plasma (pNfL) levels

Variables	Univariable exp (β)	*p* Value	Multivariable exp (β)	*p* Value
Age	1.003	0.11	Not included	—
Sex (ref = male)	0.917	0.054	Not included	—
Age at disease onset	1.004	0.057	Not included	—
Disease duration	0.9997	0.91	Not included	—
No. of previous DMTs	0.922	≤0.001	0.951	0.05
DMT just before DMT start:				
No DMT	Ref	—	Ref	—
IFN/GA	0.865	≤0.001	0.873	0.009
DMF, FGL, NTZ, RTX, TFL	0.733	≤0.001	0.784	0.002
EDSS	1.078	≤0.001	Not included	—
ARMSS	1.035	≤0.001	Not included	—
MSIS-29 physical	1.172	≤0.001	1.141	≤0.001
MSIS-29 psychological	1.078	≤0.001	Not included	—
SDMT	0.989	≤0.001	0.991	≤0.001
No. of relapses year before	1.199	≤0.001	1.134	≤0.001
Relapses year before (yes/no)	1.304	≤0.001	Not included	—
Relapses 3 months before (yes/no)	1.344	≤0.001	Not included	—

Abbreviations: ARMSS = Age-Related MS Severity Score; DMF = dimethyl fumarate; DMT = disease-modifying therapies; EDSS = Expanded Disability Status Scale; FGL = fingolimod; GA = glatiramer acetate; IFN = interferon; MSIS-29 = MS Impact Score; NTZ = natalizumab; RTX = rituximab; SDMT = Symbol Digit Modalities Test; TFL = teriflunomide.
Univariable estimates are given for all tested variables, while multivariable estimates are only provided for the variables used in the final model. The estimates β were obtained on a log scale and were back transformed (i.e., exp[β]) for ease of interpretation. Hence, exp(β) of 1.10 means an increase of 10% in the pNfL level, while 0.90 for exp(β) means a 10% decrease in the pNfL level.

(i.e., change in log-pNfL$_{N40}$) as the dependent variable and the DMTs as explanatory variables are presented in table 3. The estimates (β) were back-transformed to the original scale (exp[β]), so that, for example, a value of 0.80 translates into a 20% reduction of the baseline pNfL$_{N40}$ level. The mean change was affected by the type of DMT, with the largest mean reduction for ALM in both analyses (54% reduction [95% CI 43%–62% reduction] and 48% reduction [49%–56% reduction] respectively for the unweighted and weighted analyses) and the smallest change for TFL, for which the significance level of 0.05 was not reached (12% increase [3% reduction to 29% increase] and 7% reduction [16% reduction to 4% increase] respectively for the unweighted and weighted analyses). A post hoc analysis highlighted similarities and differences between DMT groups; the mean delta between DMF and FGL and between NTZ and ALM were not statistically different for the unweighted model. In the weighted model, the mean delta of NTZ did not differ significantly from DMF and FGL (data not shown). To remove any residual confounding, we further adjusted our model with several baseline covariates. While this dramatically increased the percentage of the variance explained, it did not change the

pattern observed with our first (weighted) model. The estimates were only slightly modified when including the log-pNfL$_{N40}$ at baseline in the model (table 4). Similar limited changes also occurred with inclusion of additional baseline covariates or stratification on PS quintiles (instead of weighting) (table 4). In order to explore the effect of previous DMTs, we further stratified on previous treatment and on baseline pNfL$_{N40}$ level (data not shown). This provided additional insights without modifying our previous observations. Finally, we also observed that the changes in log-pNfL$_{N40}$ values, EDSS, ARMSS, and MSIS-29 were all significantly correlated to each other, though often with low correlation coefficients (i.e., around 0.3 or below).

On-treatment log-pNfL$_{N40}$ levels
The analysis of the log-pNfL$_{N40}$ on treatment with either a weighted linear model (without RTX group) or with an unweighted model showed that all DMT groups had on average lower values than TFL (table 5). Adjusting for the baseline log-pNfL$_{N40}$ improved the model substantially, increasing the percentage of the explained variance from 21% to 40%, but did not affect overall estimates. Additional

Figure 2 Unweighted and weighted baseline log–neurofilament light in plasma (pNfL)$_{N40}$
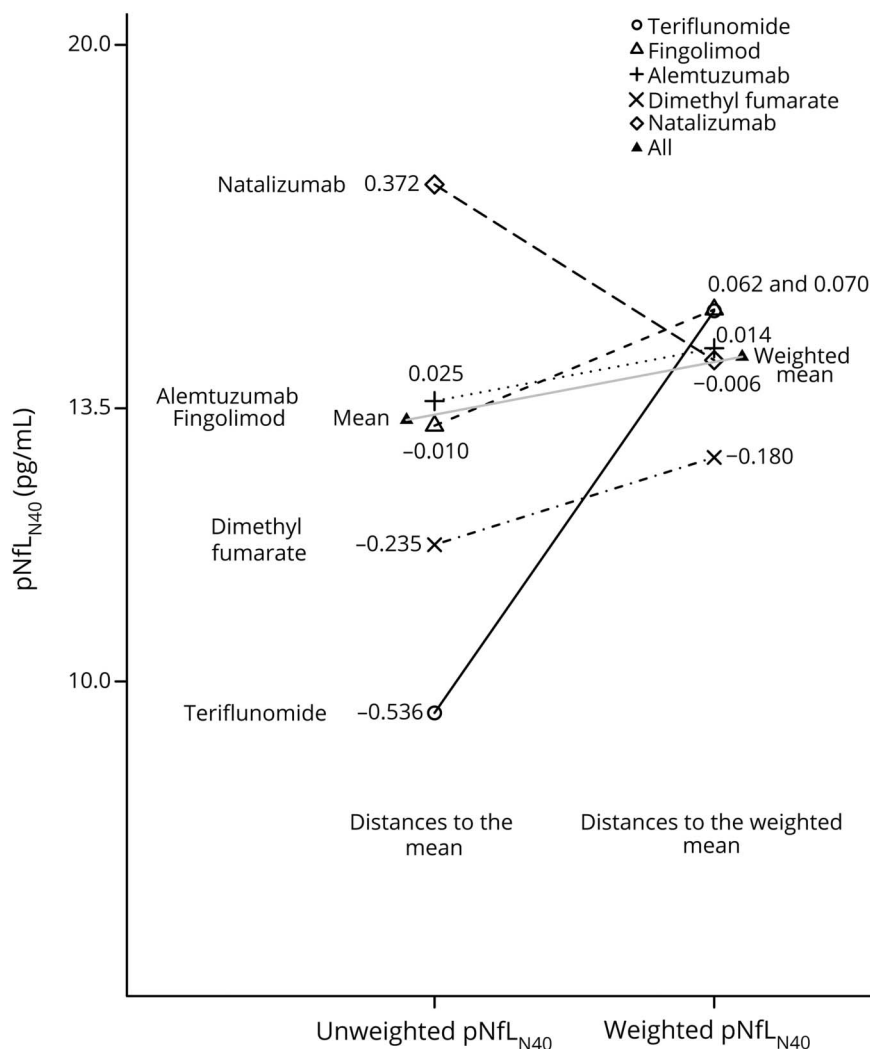
The ability of the propensity scores to correct the imbalance between disease-modifying therapy groups is shown graphically and numerically for log-pNfL$_{N40}$ levels. The distances are standardized (i.e., they do not depend on the unit in which the variable was measured). The effect of the propensity score is to decrease the standardized distances, where a standardized distance larger than 0.20 can be considered as evidence of imbalance and a potential source of bias.[21] Here, there is some small residual imbalance for dimethyl fumarate. The average of the absolute standardized distances was 0.24 before weighting and 0.05 after weighting.

adjustments did not substantially modify these estimates further. Treatment duration was tested but did not have a significant contribution.

Discussion

Disease pathogenesis in RRMS evolves over years and the availability of a growing number of treatment options creates a need for additional means to assess disease activity and response to treatment, including body fluid biomarkers.[8,23] In addition, real-world studies conducted in unselected patient populations can provide important information on questions that cannot be addressed with existing data from randomized controlled trials.[24] Along these lines, we explored how pNfL concentrations were distributed in patients with RRMS starting newer DMTs, how this distribution was associated with clinical measures and patient characteristics, and how pNfL concentrations evolved under treatment. Strengths of the study include the possibility to simultaneously compare

across multiple treatments in nonrestricted patient groups, but this approach also entails major challenges in balancing out differences in baseline characteristics, since DMT selection is heavily influenced by clinical disease characteristics. Nevertheless, by modeling on relevant variables, we demonstrate that the reduction in pNfL concentrations differs across DMTs, with the largest reduction for ALM and the smallest for TFL. This result is largely in agreement with the perceived effectiveness of the studied DMTs. Still, reductions in pNfL with DMF, FGL, and NTZ were similar even if NTZ generally is considered to have a superior effect on relapses and focal MRI lesions of the 3. This observation may be partly explained by indication bias (i.e., patients with more active disease are started on highly effective drugs); however, an interesting feature with pNfL is that it reflects both diffuse and focal neuroaxonal damage, where it may be speculated whether different DMTs affect these 2 aspects differently, for example based on their capacity to penetrate into the CNS. This will need longer follow-up studies that also integrate

Figure 3 Baseline and on-treatment mean neurofilament light in plasma (pNfL)$_{N40}$ levels in the disease-modifying therapy groups
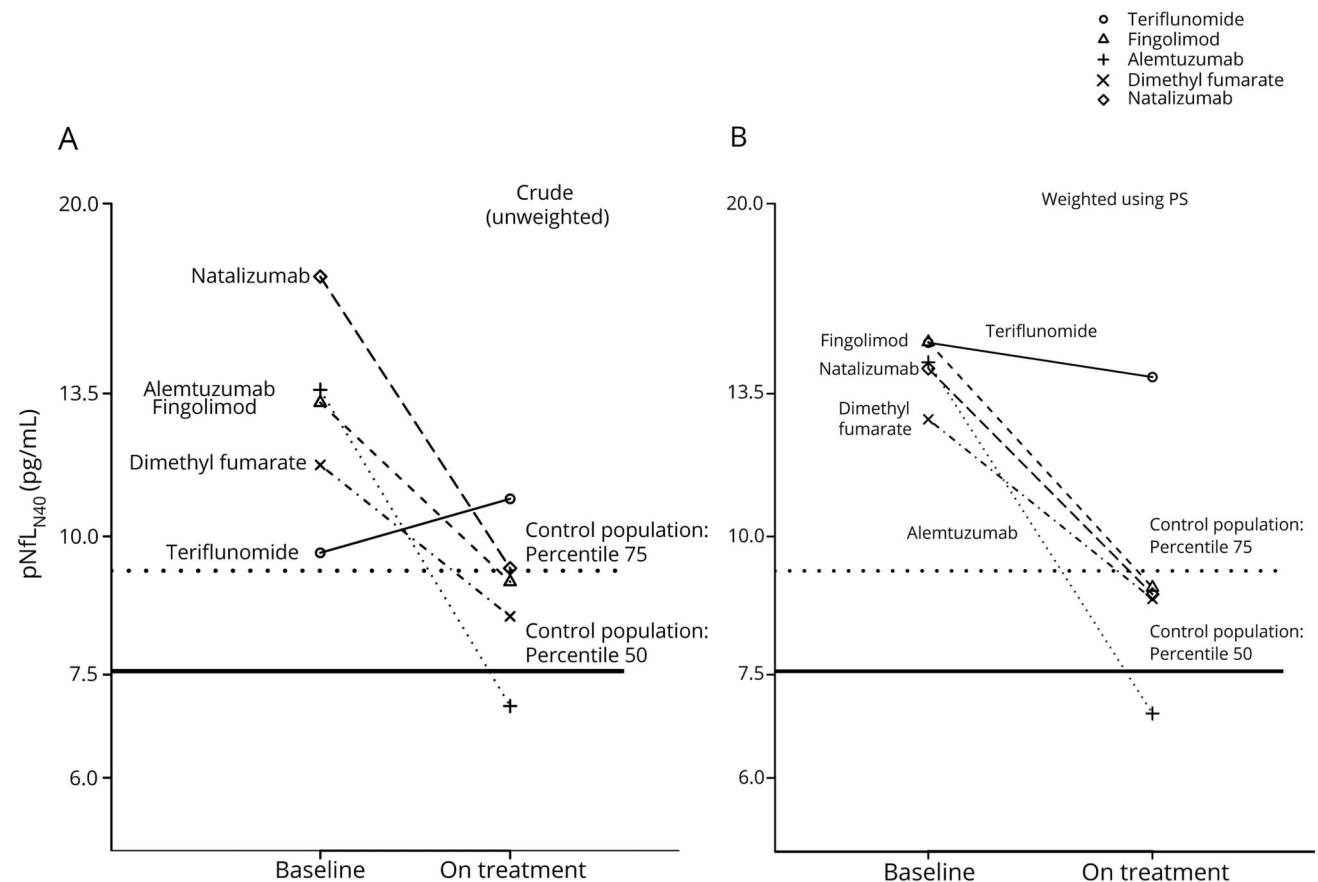
(A) Crude mean pNfL levels at baseline and on treatment. (B) Weighted mean pNfL levels at baseline and on treatment. The weights are the inverse of the propensity scores.

quantitative MRI measures. Also, the kinetics of how pNfL is affected might differ across DMTs, necessitating longer follow-up with repeated sampling. Finally, comorbidities affecting the peripheral nervous system or CNS may act as confounders. For example, leflunomide, which is related to TFL, has been shown to affect the peripheral nervous system.[25] An additional important finding is that we show how essential the baseline pNfL concentration is for correctly predicting the pNfL concentration on treatment. In fact, the percentage of the variance explained by the baseline concentration (>20%) outsized all other factors. Accordingly, inclusion of the baseline pNfL value affected estimates, increasing the differences in pNfL concentrations between the DMTs. We also find that reductions in pNfL concentrations correlated with improvements in clinical variables, such as EDSS, MSSS, and MSIS-29, though correlation coefficients were low (between 0.10 and 0.30), replicating earlier findings.[13,14] Importantly, as shown by recent studies, pNfL concentrations at diagnosis also predict important long-term outcomes, such as brain atrophy and risk to achieve clinical disability milestones.[15,26]

Whereas our data reveal differences in pNfL dynamics across the studied DMTs, we cannot rule out that differences had been achieved with a more complete model for the PS, even if our additional adjustments did not lead to major changes in the estimates. Notably, however, we did not have access to sufficiently precise MRI data, which are known to affect pNfL.[13] A further weakness is imprecise information on some measures, e.g., the lack of coding for switching from NTZ due to positive JC virus serology in the Swedish MS registry. On the other hand, the high general validity of data entered into the Swedish MS registry regarding treatment episodes and relapses was recently confirmed by a large-scale national validation against medical records.[27] Furthermore, most patients in the RTX group lacked a baseline sample, which meant that this group was excluded from analyses involving PS and that other analyses including baseline log-pNfL$_{N40}$ became less precise. Also the proportion of patients missing information for some variables that were used in the adjustment (or in the PS estimation) could also have hampered the power of our study. The observational design of the study implies that patients were not randomized to

Table 3 Mean changes in log–neurofilament light in plasma (pNfL)$_{N40}$ values between baseline and on-treatment measures provided by the unweighted and weighted analyses

	Delta: Changes in pNfL$_{N40}$ values unweighted analysis, Exp(β) (95% CI)	p Value	Delta: Changes in log-pNfL values weighted analysis, Exp(β) (95% CI)	p Value
TFL (reference)	1.119 (0.970–1.291)	Ref	0.931 (0.840–1.044)	Ref
DMF	0.652 (0.547–0.777)	≤0.001	0.739 (0.628–0.870)	≤0.001
FGL	0.615 (0.516–0.734)	≤0.001	0.644 (0.547–0.758)	≤0.001
NTZ	0.487 (0.410–0.578)	≤0.001	0.671 (0.570–0.791)	≤0.001
ALM	0.462 (0.375–0.570)	≤0.001	0.517 (0.440–0.608)	≤0.001

Abbreviations: ALM = alemtuzumab; CI = confidence interval; DMF = dimethyl fumarate; FGL = fingolimod; NTZ = natalizumab; TFL = teriflunomide.
All estimates are calculated with a linear model with the disease-modifying therapies (DMTs) as the only explanatory variable. The estimates have been back transformed to the original scale. Hence a value of 0.65 means that the on-treatment pNfL$_{N40}$ value is 0.65 times the baseline value. The estimates in column 2 were obtained from an unweighted model while the estimates in column 4 were obtained from the weighted analysis where the data were weighted by the inverse of the propensity scores. With the exception of TFL, all drugs were associated with statistically significant mean reductions. TFL was non-significantly associated with either an increase (unweighted model) or a reduction (weighted model) of the pNfL$_{N40}$ level. The significance levels indicate how the DMT groups differ from TFL.

treatment, nor were they randomly selected within the IMSE cohorts, and therefore some selection bias could have occurred. It is therefore important to relate these findings to studies exploring pNfL concentrations in the context of randomized control trials, even if such studies rarely include more than 2 DMTs.[28] As a final note, the extent different DMTs affected pNfL largely mimic their effect on the long-term risk to convert to a secondary progressive disease course, as observed in a large recent real-world study.[4] The implementation of soluble but also novel imaging biomarkers that can complement current clinical and imaging monitoring likely will lead to an increased use of more effective DMTs and reduce the risks for patients to be exposed to insufficient treatment responses, in turn improving important long-term clinical outcomes.[12,29]

We demonstrate that dynamics of pNfL are significantly influenced by specific DMTs and that the degree of pNfL reduction is correlated to clinical and patient-reported outcomes, but also that the baseline pNfL concentration exerts an unproportioned effect on on-treatment values in the medium term. In order to understand if pNfL can be used as a drug response biomarker at the individual level, further studies are needed to address the correlation of pNfL changes to long-term clinical outcomes with different DMTs, as well as if modeling of pNfL dynamics can be improved further by including additional variables such as MRI data or more frequent measurements.

Study funding

Table 4 Mean changes in log–neurofilament light in plasma (pNfL)$_{N40}$ values between baseline and on-treatment measures provided by the analyses using propensity scores: weighting by their inverse and adjusting for baseline log-pNfL$_{N40}$ and additional baseline variables, or stratifying on quintile of the propensity score distribution

	Weighted model adjusted for baseline pNfL$_{N40}$	p Value	Weighted model adjusted for baseline pNfL$_{N40}$ and additional variables	p Value	Stratified model (not adjusted)	p Value
TFL (reference)	1.01 (0.94–1.09)	Ref	0.86 (0.73–1.01)	Ref	0.82 (0.64–1.04)	Ref
DMF	0.66 (0.60–0.73)	≤0.001	0.68 (0.62–0.75)	≤0.001	0.66 (0.56–0.79)	≤0.001
FGL	0.65 (0.58–0.71)	≤0.001	0.67 (0.61–0.74)	≤0.001	0.57 (0.48–0.68)	≤0.001
NTZ	0.65 (0.59–0.72)	≤0.001	0.66 (0.60–0.73)	≤0.001	0.65 (0.55–0.78)	≤0.001
ALM	0.50 (0.46–0.56)	≤0.001	0.52 (0.47–0.57)	≤0.001	0.54 (0.46–0.63)	≤0.001

Abbreviations: ALM = alemtuzumab; DMF = dimethyl fumarate; FGL = fingolimod; NTZ = natalizumab; TFL = teriflunomide.
All estimates were calculated with a linear model with the disease-modifying therapies (DMTs) as the main explanatory variable and using propensity scores: weighting by their inverse and adjusting for baseline log-pNfL$_{N40}$ (column 2), weighting by their inverse and adjusting for baseline log-pNfL$_{N40}$, sex, age, Expanded Disability Status Scale, Age-Related MS Severity Score, Symbol Digit Modalities Test, age at disease onset, disease duration, and treatment status just before DMT start (column 4), or stratifying on the quintile of the propensity score distribution without adjustment (column 6). The estimates have been back transformed to the original scale. Hence a value of 0.65 means that the on-treatment pNfL$_{N40}$ value is 0.65 times the baseline value. The significance levels indicate how the DMT groups differ from TFL.

Table 5 Comparison of on-treatment log–neurofilament light in plasma $(pNfL)_{N40}$ between treatment groups with several statistical linear models, unweighted and adjusted or weighted without and with adjustment

	Unweighted				Weighted (propensity score)			
	(1)	**p Value**	**(2)**	**p Value**	**(3)**	**p Value**	**(4)**	**p Value**
TFL (ref)	13.9	Ref	13.9	Ref	13.9	Ref	13.9	Ref
DMF	10.5	≤0.001	11.1	≤0.001	8.8	≤0.001	9.6	≤0.001
FGL	11.0	≤0.001	11.9	0.012	9.0	≤0.001	9.5	≤0.001
NTZ	10.5	≤0.001	12.3	0.042	8.9	≤0.001	9.4	≤0.001
ALM	8.2	≤0.001	8.7	≤0.001	6.9	≤0.001	7.6	≤0.001
RTX	11.9	0.4	13.7	0.8	Group not included			
%	24%		8%		21%		46%[a]	

Abbreviations: ALM = alemtuzumab; DMF = dimethyl fumarate; FGL = fingolimod; NTZ = natalizumab; RTX = rituximab; TFL = teriflunomide.
% = Percentage of the variance explained (i.e., r^2 of the model). (1) $pNfL_{N40}$ value on treatment (pg/mL): unweighted adjusted model; (2) same as (1) without including the baseline $log\text{-}pNfL_{N40}$; (3) $pNfL_{N40}$ value on treatment (pg/mL): weighted model without adjustment; (4) same as (3) with the same adjustment as (1). The on-treatment $log\text{-}pNfL_{N40}$ value in the disease-modifying therapy (DMT) groups were calculated in reference to TFL (for which the on-treatment weighted mean pNfL value was 13.9 pg/mL). (1) The unweighted model adjusted for several baseline covariates (baseline $log\text{-}pNfL_{N40}$, age at DMT start, sex, age at disease onset, disease duration, and treatment status just before DMT start) provided estimates for on-treatment $log\ pNfL_{N40}$, which were lower for all DMTs compared to TFL, though not significantly for RTX. (2) The differences with TFL were attenuated when removing the baseline $log\text{-}pNfL_{N40}$ from the model. (3 and 4) In contrast, the differences with TFL were exacerbated in the weighted models. Adjusting or not for covariates in these models slightly modified the estimates. Note that RTX was not included in the weighted analyses. The significance levels indicate whether the values are significantly different from the reference (TFL).
[a] 21% for DMT; 19% for baseline $log\text{-}pNfL_{N40}$; 6% for all other covariates.

Disclosure

Publication history

Appendix Authors

Name	Location	Role	Contribution
Bénédicte Delcoigne, PhD	Karolinska Institutet, Stockholm, Sweden	Author	Conceived and designed the study, performed the data management and the statistical analyses, interpreted the results, drafted the manuscript
Ali Manouchehrinia, PhD	Karolinska Institutet, Stockholm, Sweden	Author	Interpreted the results, critically reviewed the manuscript
Christian Barro, MD	University Hospital Basel, Switzerland	Author	Performed the data acquisition, interpreted the results, critically reviewed the manuscript
Pascal Benkert, PhD	University Hospital Basel, Switzerland	Author	Interpreted the results, critically reviewed the manuscript
Zuzanna Michalak, PhD	University Hospital Basel, Switzerland	Author	Performed the data acquisition, interpreted the results, critically reviewed the manuscript
Ludwig Kappos, MD	University Hospital Basel, Switzerland	Author	Interpreted the results, critically reviewed the manuscript
David Leppert	University Hospital Basel, Switzerland	Author	Interpreted the results, critically reviewed the manuscript
Jon A. Tsai, MD	Sanofi Genzyme, Sweden	Author	Interpreted the results, critically reviewed the manuscript

Appendix *(continued)*

Name	Location	Role	Contribution
Tatiana Plavina, PhD	Biogen, Cambridge, MA	Author	Interpreted the results, critically reviewed the manuscript
Bernd C. Kieseier, MD	Biogen, Cambridge, MA	Author	Interpreted the results, critically reviewed the manuscript
Jan Lycke, MD	University of Gothenburg, Sweden	Author	Interpreted the results, critically reviewed the manuscript
Lars Alfredsson, PhD	Karolinska Institutet, Stockholm, Sweden	Author	Interpreted the results, critically reviewed the manuscript
Ingrid Kockum, PhD	Karolinska Institutet, Stockholm, Sweden	Author	Performed the data acquisition and the initial data management, interpreted the results, critically reviewed the manuscript
Jens Kuhle, MD	University Hospital Basel, Switzerland	Author	Performed the data acquisition, interpreted the results, critically reviewed the manuscript
Tomas Olsson	Karolinska Institutet, Stockholm, Sweden	Author	Conceived and designed the study, interpreted the results, critically reviewed the manuscript
Fredrik Piehl, MD	Karolinska Institutet, Stockholm, Sweden	Author	Conceived and designed the study, interpreted the results, drafted the manuscript

References

1. Filippi M, Bar-Or A, Piehl F, et al. Multiple sclerosis. Nat Rev Dis Primers 2018;4:43.
2. Tallantyre EC, Bo L, Al-Rawashdeh O, et al. Clinico-pathological evidence that axonal loss underlies disability in progressive multiple sclerosis. Mult Scler 2010; 16:406–411.
3. Trapp BD, Stys PK. Virtual hypoxia and chronic necrosis of demyelinated axons in multiple sclerosis. Lancet Neurol 2009;8:280–291.
4. Brown JWL, Coles A, Horakova D, et al. Association of initial disease-modifying therapy with later conversion to secondary progressive multiple sclerosis. JAMA 2019; 321:175–187.
5. Steenwijk MD, Geurts JJ, Daams M, et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain 2016;139:115–126.
6. Sormani MP, Arnold DL, De Stefano N. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. Ann Neurol 2014;75:43–49.
7. Eshaghi A, Prados F, Brownlee WJ, et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. Ann Neurol 2018;83:210–222.
8. Khalil M, Teunissen CE, Otto M, et al. Neurofilaments as biomarkers in neurological disorders. Nat Rev Neurol 2018;14:577–589.
9. Gunnarsson M, Malmestrom C, Axelsson M, et al. Axonal damage in relapsing multiple sclerosis is markedly reduced by natalizumab. Ann Neurol 2011;69:83–89.
10. Kuhle J, Disanto G, Lorscheider J, et al. Fingolimod and CSF neurofilament light chain levels in relapsing-remitting multiple sclerosis. Neurology 2015;84:1639–1643.
11. Novakova L, Axelsson M, Khademi M, et al. Cerebrospinal fluid biomarkers as a measure of disease activity and treatment efficacy in relapsing-remitting multiple sclerosis. J Neurochem 2017;141:296–304.
12. de Flon P, Gunnarsson M, Laurell K, et al. Reduced inflammation in relapsing-remitting multiple sclerosis after therapy switch to rituximab. Neurology 2016;87: 141–147.
13. 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: 2382–2391.

14. Disanto G, Barro C, Benkert P, et al. Serum neurofilament light: a biomarker of neuronal damage in multiple sclerosis. Ann Neurol 2017;81:857–870.

15. Manouchehrinia A, Stridh P, Khademi M, et al. Plasma neurofilament light levels are associated with risk of developing sustained disability in multiple sclerosis. ECTRIMS Online 2018;228223:P378.

16. Novakova L, Zetterberg H, Sundstrom P, et al. Monitoring disease activity in multiple sclerosis using serum neurofilament light protein. Neurology 2017;89:2230–2237.

17. Manouchehrinia A, Westerlind H, Kingwell E, et al. Age related multiple sclerosis severity score: disability ranked by age. Mult Scler 2017;23:1938–1946.

18. Zhang Z. Variable selection with stepwise and best subset approaches. Ann Transl Med 2016;4:136.

19. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011;46:399–424.

20. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 2015;34:3661–3679.

21. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med 2013;32:3388–3414.

22. Nguyen TL, Collins GS, Spence J, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. BMC Med Res Methodol 2017;17:78.

23. Comabella M, Montalban X. Body fluid biomarkers in multiple sclerosis. Lancet Neurol 2014;13:113–126.

24. Trojano M, Tintore M, Montalban X, et al. Treatment decisions in multiple sclerosis: insights from real-world observational studies. Nat Rev Neurol 2017;13:105–118.

25. Kim HK, Park SB, Park JW, et al. The effect of leflunomide on cold and vibratory sensation in patients with rheumatoid arthritis. Ann Rehabil Med 2012;36:207–212.

26. 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 Epub 2019 Aug 12.

27. Alping P, Piehl F, Langer-Gould A; on behalf of the COMBAT-MS Study Group, et al. Validation of the Swedish multiple sclerosis register: further improving a resource for pharmacoepidemiologic evaluations. Epidemiology 2019;30:230–233.

28. Kuhle J, Kropshofer H, Haering DA, et al. Blood neurofilament light chain as a biomarker of MS disease activity and treatment response. Neurology 2019;92:e1007–e1015.

29. Granqvist M, Boremalm M, Poorghobad A, et al. Comparative effectiveness of rituximab and other initial treatment choices for multiple sclerosis. JAMA Neurol 2018;75:320–327.

Neurology®

Blood neurofilament light levels segregate treatment effects in multiple sclerosis

Bénédicte Delcoigne, Ali Manouchehrinia, Christian Barro, et al.

This information is current as of February 11, 2020

Updated Information & Services	including high resolution figures, can be found at: http://n.neurology.org/content/early/2020/02/11/WNL.0000000000009097.full
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): **Cohort studies** http://n.neurology.org/cgi/collection/cohort_studies **Multiple sclerosis** http://n.neurology.org/cgi/collection/multiple_sclerosis
Permissions & Licensing	Information about reproducing this article in parts (figures,tables) or in its entirety can be found online at: http://www.neurology.org/about/about_the_journal#permissions
Reprints	Information about ordering reprints can be found online: http://n.neurology.org/subscribers/advertise

AMERICAN ACADEMY OF
NEUROLOGY®