ARTICLE OPEN ACCESS CLASS OF EVIDENCE

Long-term prognostic value of longitudinal measurements of blood neurofilament levels

Dieter A. Häring, PhD, Harald Kropshofer, PhD, Ludwig Kappos, MD, Jeffrey A. Cohen, MD, Anuja Shah, PhD, Rolf Meinert, PhD, David Leppert, MD, Davorka Tomic, DVM, PhD, and Jens Kuhle, MD, PhD

Correspondence
Dr. Kuhle
Jens.Kuhle@usb.ch

Neurol Neuroimmunol Neuroinflamm 2020;7:e856. doi:10.1212/NXI.0000000000000856

Abstract

Objective

To assess the long-term prognostic value of an integral of longitudinal measurements of plasma neurofilament light chain levels (NfL_{long}) over 12 and 24 months vs single neurofilament light chain (NfL) measurements in patients with relapsing-remitting MS (RRMS) and its additional value when combined with clinical and MRI measures.

Methods

This analysis included continuously fingolimod-treated patients with RRMS from the 24-month FTY720 Research Evaluating Effects of Daily Oral therapy in Multiple Sclerosis (FREEDOMS)/12-month Trial Assessing Injectable Interferon vs FTY720 Oral in Relapsing–Remitting Multiple Sclerosis (TRANSFORMS) phase 3 trials and their long-term extension, LONGTERMS. Patients were classified into high (≥30 pg/mL, n = 110) and low (<30 pg/mL, n = 164) NfL categories based on the baseline (BL) NfL value or the geometric mean NfL_{long} calculated over 12 and 24 months to predict disability-related outcomes and brain volume loss (BVL). The additional prognostic value of NfL was quantified using the area under the receiver operating characteristic (ROC) curve.

Results

A single high (vs low) NfL measure at BL was prognostic of a higher risk of reaching Expanded Disability Status Scale (EDSS) score ≥4 earlier (hazard ratio [HR] = 2.19; 95% CI = 1.21–3.97) and higher BVL over 120 months (difference: −1.12%; 95% CI = −2.07 to −0.17). When NfL_{long} was measured over 24 months, high NfL was associated with a higher risk of reaching EDSS score ≥4 (HR = 7.91; 95% CI = 2.99–20.92), accelerated 6-month confirmed disability worsening (HR = 3.14; 95% CI = 1.38–7.11), and 20% worsening in the Timed 25-Foot Walk Test (HR = 3.05; 95% CI = 1.38–6.70). Area under the ROC curve was consistently highest in models combining NfL with clinical and MRI measures.

Conclusions

NfL_{long} had a higher prognostic value than single NfL assessments on long-term outcomes in RRMS. Combining it with clinical and MRI measures increased sensitivity and specificity to predict long-term disease outcomes.

Classification of evidence

This study provides Class I evidence that NfL_{long} was more strongly associated with long-term outcomes than single NfL assessments in patients with RRMS.

MORE ONLINE

→ **Class of Evidence**
Criteria for rating therapeutic and diagnostic studies
NPub.org/coe

From the Novartis Pharma AG (D.A.H., H.K., D.T.), Basel; Research Center for Clinical Neuroimmunology and Neuroscience Basel (L.K., D.L., J.K.), Departments of Medicine, Clinical Research, Biomedicine and Biomedical Engineering, University Hospital and University of Basel, Switzerland; Department of Neurology (J.A.C.), Mellen MS Center, Neurological Institute, Cleveland Clinic, OH; Novartis Healthcare Pvt. Ltd. (A.S.), Hyderabad, India; and DATAMAP GmbH (R.M.), Freiburg, Germany.

Go to Neurology.org/NN for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by Novartis Pharma AG, Basel, Switzerland.

Glossary

ARBA = annualized rate of brain atrophy; **AUC** = area under the curve; **BVL** = brain volume loss; **BL** = baseline; **CM** = clinical model; **EDSS** = Expanded Disability Status Scale; **Gd+** = gadolinium enhancing; **HR** = hazard ratio; **NfL** = neurofilament light chain; **PBVC** = percentage brain volume change; **ROC** = receiver operating characteristic; **RRMS** = relapsing-remitting MS; **PASAT** = Paced Auditory Serial Addition Test; **T25FWT** = Timed 25-Foot Walk Test; **9HPT** = 9-Hole Peg Test; **6m-CDW** = 6-month confirmed disability worsening.

MS is a chronic autoimmune disorder, characterized by CNS inflammation and neurodegeneration, leading to accumulation of disability.[1] The clinical disease course of MS is heterogeneous and remains a challenge for prognosis and therapeutic decision making in individual patients based on clinical and MRI measures.[2,3]

Neurofilament light chain (NfL) is a cytoskeletal protein exclusively expressed by neurons[4,5]; its release into the CSF and blood[4,6] is a highly specific sign of neuronal injury. The strong correlation of CSF and serum/plasma levels of NfL has allowed its establishment as a blood biomarker for monitoring disease activity and treatment response.[6–13] Furthermore, single measurements of elevated NfL concentrations at baseline (BL) are associated with on-study relapses, MRI lesions, brain volume loss (BVL), spinal cord atrophy, and disability worsening.[7–9] However, the prognostic value of NfL related to long-term disability outcomes, and particularly the added value when combined with clinical and MRI markers, has so far not been explored in the long-term follow-up of phase 3 studies.

We hypothesized that an integral of longitudinal measurements of NfL (NfL_{long}) over 12 or 24 months would have superior prognostic value for long-term outcomes over single (i.e., BL) NfL measures in relapsing-remitting MS (RRMS). The present analysis of data from 2 phase 3 clinical studies and their extensions aimed to quantify the long-term prognostic value of an integral of NfL_{long} over 12 or 24 months in patients with RRMS under fingolimod (Gilenya; Novartis Pharma AG, Basel, Switzerland) therapy for disability worsening over a 10-year follow-up. Furthermore, we assessed whether NfL provides additional value when combined with conventional clinical and MRI measures to improve long-term prognosis of disability outcomes and BVL in patients with MS.

Methods

Study design and patient population

The present post hoc biomarker analysis included pooled data from patients with RRMS who were randomly assigned to receive fingolimod 0.5 mg once daily during the core period of the 24-month FTY720 Research Evaluating Effects of Daily Oral therapy in Multiple Sclerosis (FREEDOMS) (NCT00289978) [14] or 12-month Trial Assessing Injectable Interferon vs FTY720 Oral in Relapsing -Remitting Multiple Sclerosis (TRANSFORMS) (NCT00340834)[15] phase 3 trials (both trials had essentially the same inclusion/exclusion criteria), continued on the same treatment and dose in the respective extension studies (NCT00662649 and NCT00340834),[16,17] and thereafter transitioned into the open-label long-term extension LONGTERMS study for up to 10 years (NCT01201356).[18] Details of the individual study design and patient population are reported elsewhere.[14–18] NfL analysis was performed in all samples where informed consent was granted, irrespective of clinical outcomes.

NfL assessments

NfL was analyzed using a single molecule array (SIMOA) immunoassay (Quanterix Corporation, Billerica, MA) in all patients who gave consent for an exploratory analysis of their stored ethylenediaminetetraacetic acid–treated plasma samples.[8,10] Plasma samples of NfL were collected during the core study period at BL and at months 6, 12, and 18, and 24 (FREEDOMS only) and analyzed later at the University Hospital, Basel, Switzerland. Laboratory personnel were blinded to treatment allocation with no access to clinical data. The biostatistical analyses were performed at DATAMAP GmbH, Freiburg, Germany.

Outcome measures

The prognostic value of NfL was tested separately for percentage brain volume change (PBVC), time to Expanded Disability Status Scale (EDSS) score ≥4.0, time to first 6-month confirmed disability worsening (6m-CDW), time to 6-month confirmed 20% worsening in the Timed 25-Foot Walk Test (T25FWT), time to 6-month confirmed 20% worsening in the 9-Hole Peg Test (9HPT), and time to 6-month confirmed 20% worsening in the Paced Auditory Serial Addition Test (PASAT).

The prognostic value of NfL, clinical measures with/without MRI, or NfL in combination with clinical and MRI measures was measured for all the long-term outcomes up to month 84 based on the following combinations of different predictor sets: clinical model (CM), CM plus MRI predictor set (CM + MRI), CM plus NfL predictor set (CM + NfL), and CM plus MRI predictor set and NfL predictor set (CM + MRI + NfL).

Standardized MRI scans were obtained at the screening visit and at months 6, 12, and 24 (FREEDOMS only) during the core phase and yearly in the extension phase. Brain volume change was measured using structural image evaluation using normalization of the atrophy (SIENA; v3.3 [TRANSFORMS], and v4.2 [FREEDOMS]) software (FMRIB [Oxford Centre for Functional Magnetic Resonance Imaging of the Brain], Oxford UK) using the provider's default settings (in all cases, the MS MRI lesions were not masked in the process). The

annualized rate of brain atrophy (ARBA) was calculated from the PBVC, as = ([PBVC/100 + 1][365.25/#days)] −1) × 100 where "days" stands for the scan date relative to day 1 for the primary analysis and relative to the date of the month 6 scan for the sensitivity analysis.

EDSS scores were determined every 3 months. T25FWT, 9HPT, and PASAT scores were measured every 6 months in the core phase and yearly in the extension phase.[14,15,17,19]

Level of evidence

This post hoc analysis provides level I evidence for long-term prognostic value of an integral of NfL_{long} over 24 months to determine disability worsening in patients with RRMS over 10 years, using data from phase 3 fingolimod studies and their extensions.

Statistical analysis

The present analysis classified patients into high (≥30 pg/mL) or low (<30 pg/mL) NfL level categories,[10] based on 3 classifications of NfL as follows: (1) a single measurement at BL (before study treatment initiation; NfL [BL]), (2) the geometric mean over 1 year (2–3 values per patient at BL, month 6, and month 12; NfL [BL-month 12]), and (3) the geometric mean over 2 years (3–5 values per patient at BL, months 6, 12, 18, and 24—at least 1 value from month 18 or 24; NfL [BL-month 24]). Patients without a BL NfL assessment still could contribute to the integral measurements over 12 and/or 24 months. The analysis was performed in all patients who received fingolimod during the respective studies and remained on fingolimod in the extension study (patients who discontinued from fingolimod and switched to other disease-modifying therapies had to discontinue from the study and were censored at this time point). All patients who had at least 1 NfL assessment (at BL) and the respective demographic and disease characteristic data could contribute to the analysis. Only events that occurred post-BL, or after the interval used for the categorization of patients by NfL levels, were counted in the statistical analysis. When patients were categorized by BL NfL, all post-BL outcome events were considered; when patients were categorized by the geometric mean NfL level in the first (or second) year, only outcome events with an onset after the first (or second) year were included in the statistical analysis.

The prognostic value of NfL for patients reaching EDSS score ≥4.0, 6m-CDW, and 20% worsening on the T25FWT, 9HPT, or PASAT was analyzed using the log-rank test and the Cox proportional hazards model with adjustments for sex, age, disease duration, number of relapses in the 2 years before the study, a reference value of the respective analysis outcome (EDSS, T25FWT, 9HPT, or PASAT) according to the analysis period (BL, month 12, and month 24), and geometric mean NfL by category (high vs low) according to the analysis period. Furthermore, Kaplan-Meier plot results of time-to-event analyses are reported with p values from the log-rank test across NfL categories, with hazard ratios (HRs) and 95% CIs from the Cox proportional hazards model.

The prognostic value of NfL for PBVC was analyzed using a linear mixed model for repeated measures with adjustments for sex, age, duration of MS, number of relapses in the 2 years before the study, BL NBV, gadolinium-enhancing (Gd+) T1 lesion number at the beginning of the analysis period, T2 lesion volume at the beginning of the analysis period, and geometric mean NfL by category (high vs low) according to the analysis period (BL, month 12, and month 24). For analysis periods starting at month 12 or 24, the model also included change from BL to month 12 or 24 in T2 lesion volume. Furthermore, the repeated measures model included interaction terms between visit and NfL category and between visit and NBV at BL to account for the possibility that the relevance of BL assessments might decrease for PBVC observations measured post-BL.

To investigate whether NfL has additional prognostic value over clinical and MRI measures, all outcomes measured up to month 84 were dichotomized (long-term disability event: yes/no; BVL >0.4%/y: yes/no) and analyzed using logistic regression models. The CM contained the following covariates: sex, age, disease duration, number of relapses in the 2 years before the study, and a reference value (at month 12 or 24) of the respective outcome (EDSS, T25FWT, 9HPT, or PASAT score) taken at the start of the analysis period (at BL, month 12, or month 24). The MRI predictor set for analysis of the period from BL onward consisted of BL assessments of normalized brain volume (NBV), number of Gd+ T1 lesions, and T2 lesion volume. The MRI predictor set for analyses of the period from month 12 or 24 onward consisted of NBV at BL, T2 lesion volume at BL, T2 lesion volume change from BL to month 12 or 24, number of Gd+ T1 lesions at month 12 or 24, and PBVC from BL to month 12 or 24. The prognostic value of the various models was compared by the area under the receiver operating characteristic (ROC) curve. In the area under the ROC curve, the true positive rate (sensitivity) is plotted against the false positive rate (1 – specificity) across all possible cutoff values; the higher the area under the ROC curve, the better the model.[20] In the best case, the area under the ROC curve is one, corresponding to 100% sensitivity and 100% specificity; in contrast, a random classification would lead to an area under the curve (AUC) of 0.5.

Standard protocol approvals, registrations, and patient consents

The protocols and amendment of studies included in the present analysis were originally reviewed and approved by the Independent Ethics Committees and Institutional Review Boards for each center per local regulations. All patients or legally accepted representatives of patients provided written informed consent before study entry for the present analysis. The study was conducted in compliance with the ethical principles of the Declaration of Helsinki and the International Conference on Harmonisation Good Clinical Practice Guidelines.

Data availability

Anonymized data will be made available to qualified external researchers, with requests reviewed and approved by an independent review panel on the basis of scientific merit.

Results

Patient disposition and BL characteristics

Of the full FREEDOMS/TRANSFORMS analysis set of patients who received fingolimod 0.5 mg once daily during the core period, 301 patients had at least 1 NfL value, and 274 had an available BL assessment; patients without a BL NfL value could still be included in analyses on NfL measured over 12 (n = 274) or 24 (n = 132) months. The numbers of NfL values of patients contributing to the average geometric mean over 12 and 24 months are presented in table 1. Demographic and BL characteristics of patients who had an evaluable NfL assessment at BL aligned with the overall trial populations of FREEDOMS and TRANSFORMS (table 2). At BL, the geometric mean of NfL was 29.7 pg/mL (table 2), and 110 patients (37%) had high NfL levels (≥30 pg/mL).

The mean age and sex distribution of patients were similar between the high and low NfL categories (table 3). At BL, however, patients with high NfL had experienced a higher number of relapses before study entry, had higher EDSS scores, more Gd+ lesions, and higher T2 lesion volume compared with patients with low NfL. Patients with high BL NfL had higher EDSS scores at months 12 and 24 and lower PASAT scores at month 24; the loss of brain volume over the follow-up was more pronounced in high NfL patients. The percentage of patients completing months 24, 48, 84, and 96 was similar between the high and low NfL categories.

Prognosis of long-term outcomes by NfL

Disability-related outcomes

A single high (compared with low) NfL measurement at BL was associated with a 2-fold increase in the hazard of reaching EDSS ≥4.0 (HR = 2.19; 95% CI = 1.21–3.97; figure 1–1.1A), but was not predictive of the risk of reaching 6m-CDW (figure 1–1.2A), or 20% worsening in the T25FWT (figure 1–1.3A), 9HPT (figure 1–1.4A), or PASAT (figure 1–1.5A).

When using the geometric mean of NfL_{long} collected over 12 months (up to 3 measurements), a higher predictive value for reaching EDSS ≥4 was observed (HR = 2.78; 95% CI = 1.51–5.10; figure 1–1.1B). Moreover, the geometric mean of NfL_{long} collected over 12 months predicted 20% worsening in the PASAT (HR = 2.59; 95% CI = 1.04–6.47; figure 1–1.5B. However, it was not predictive of the risk of reaching 6m-CDW (HR = 1.53; 95% CI = 0.89–2.62; figure 1–1.2B) or 20% worsening in the T25FWT (figure 1–1.3B) or 9HPT (figure 1–1.4B).

A high (compared with low) geometric mean of NfL_{long} collected over 24 months (up to 5 measurements) was associated with an 8-fold increase in the hazard of reaching EDSS score ≥4 (HR = 7.91; 95% CI = 2.99–20.92; figure 1–1.1C) and a 3-fold increase in the hazard of reaching 6m-CDW (HR = 3.14; 95% CI = 1.38–7.11; figure 1–1.2C) and 20% worsening in the

Table 1 Visit patterns of patients contributing to the average geometric mean NfL values over 12 and 24 months

BL N = 274	M6 N = 260	M12 N = 269	M18 N = 122	M24 N = 130	Frequency (patients, n)
No. of NfL values of patients contributing to the average geometric mean over 12 months					
	✓	✓			19
✓		✓			16
✓	✓				11
✓	✓	✓			228
					Total: 274
No. of NfL values of patients contributing to the average geometric mean over 24 months					
		✓	✓	✓	1
	✓	✓		✓	1
	✓	✓	✓	✓	4
✓		✓	✓	✓	5
✓	✓	✓		✓	9
✓	✓	✓	✓		2
✓	✓	✓	✓	✓	110
					Total: 132

Abbreviations: BL = baseline; M = month; NfL = neurofilament light chain.
✓Indicates that the NfL assessment was available at that particular time point.

Table 2 Patient demographics and disease characteristics (total population)

Characteristics	Fingolimod 0.5 mg (NfL set) N = 301	FREEDOMS (full analysis set) N = 1,272	TRANSFORMS (full analysis set) N = 1,280
Age (y)	37.0 (30, 44)	37.0 (30, 43)	36.0 (30, 43)
Female, n (%)	198 (65.8)	889 (69.9)	861 (67.3)
Duration of MS since first symptoms (y)	6.6 (3.2, 12.4)	6.7 (3.0, 11.9)	5.9 (2.4, 10.7)
No. of relapses in the 2 y before screening	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)
Prior MS treatment, n (%)	166 (55.1)	520 (40.9)	745 (58.2)
EDSS scores at BL	2.0 (1.5, 3.5)	2.0 (1.5, 3.5)	2.0 (1.5, 3.0)
EDSS score at M12	2.0 (1.5, 3.5)	2.0 (1.5, 3.5)	2.0 (1.5, 3.0)
EDSS score at M24	2.0 (1.5, 3.5)	2.0 (1.5, 3.5)	2.0 (1.5, 3.0)
PASAT score at BL	52.0 (44.0, 57.0)	52.0 (43.0, 57.0)	52.0 (42.0, 57.0)
PASAT score at M12	53.0 (47.0, 58.0)	53.0 (44.0, 58.0)	53.0 (45.0, 57.0)
PASAT score at M24	55.0 (49.0, 58.0)	54.0 (46.0, 58.0)	54.0 (45.0, 58.0)
NBV (cm^3)	1,521.9 (1,466.3, 1,575)	1,520.4 (1,461.3, 1,574.0)	1,529.5 (1,473.5, 1,577.5)
Change in brain volume from BL to M12 (%)	−0.35 (−0.81, 0.08)	−0.40 (−1.0, 0.07)	−0.30 (−0.7, 0.1)
Change in brain volume from BL to M24 (%)	−0.60 (−1.3, −0.2)	−0.78 (−1.7, −0.2)	−0.50 (−1.2, −1.0)
Presence of Gd+ T1 lesions at BL, n (%)	121 (40.5)	480 (38.1)	437 (34.7)
Number of Gd+ T1 lesions at BL	0 (0, 1)	0 (0, 1)	0 (0, 1)
T2LV at BL (cm^3)	3.2 (1.5, 7.6)	3.4 (1.3, 8.3)	2.8 (1.1, 6.7)
Change in T2LV from BL to M12 (cm^3)	0.014 (−0.3, 0.3)	−0.002 (−0.2, 0.3)	0.059 (−0.13, 0.49)
Change in T2LV from BL to M24 (cm^3)	0.009 (−0.32, 0.37)	−0.003 (−0.23, 0.48)	0.13 (−0.12, 0.64)
Follow-up duration (y)	8.8 (3.7, 9.2)	8.5 (2.2, 9.4)	6.1 (1.7, 8.9)
Patients who completed M24, n (%)	268 (89.0)	1,127 (88.6)	983 (76.8)
Patients who completed M48, n (%)	229 (76.1)	833 (65.5)	793 (62.0)
Patients who completed M84, n (%)	199 (66.1)	697 (54.8)	633 (49.5)
Patients who completed M96, n (%)	188 (62.5)	658 (51.7)	588 (45.9)
Patients with ≥1 NfL assessment, geometric mean (pg/mL)	N = 301	N = 277	N = 473
BL	29.70	30.09	26.00
M12	17.72	21.67	17.15
M24	17.96	21.27	..
BL-M12	21.42	23.98	20.12
BL-M24	20.50	22.88	..

Abbreviations: BL = baseline; EDSS = Expanded Disability Status Scale; Gd+ = gadolinium enhancing; M = month; NBV = normalized brain volume; NfL = neurofilament light chain; PASAT = Paced Auditory Serial Addition Test; Q = quartile; T2LV = T2 lesion volume.
Summary statistics are presented as median (Q1, Q3), unless stated otherwise; 301 patients in the fingolimod 0.5 mg group had at least 1 NfL assessment, but only 274 had a BL NfL assessment.

T25FWT (HR = 3.05; 95% CI = 1.38–6.70; figure 1–1.3C). However, it was not predictive of reaching 20% worsening on the 9HPT (figure 1–1.4C) or PASAT (figure 1–1.5C) in this data set.

Change in brain volume

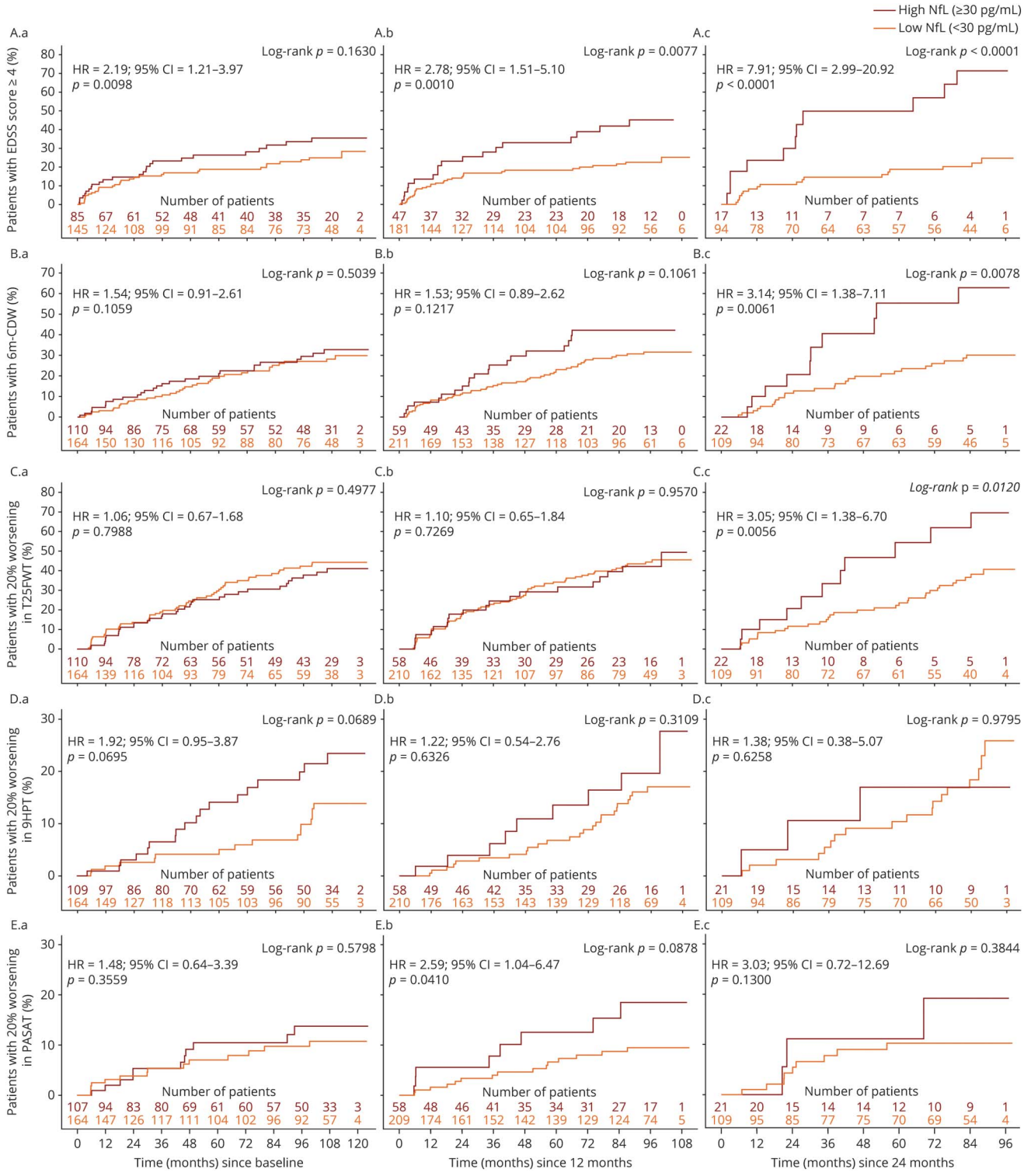
Patients with high (compared with low) NfL levels at BL lost more brain volume over 120 months (least square mean difference between the high and low category: −1.12%; 95% CI =

Table 3 Patient demographics and disease characteristics at BL, M12, and M24 (by NfL category at BL)

	NfL category		
	<30 pg/mL	≥30 pg/mL	
	n = 164	n = 110	*p* Value
Age (y)	37.0 (31.0, 44.5)	35.5 (29.0, 43.0)	NS
Female, n (%)	111 (67.7)	73 (66.4)	NS
MS duration since first symptoms (y)	7.2 (3.2, 13.2)	5.8 (3.2, 10.1)	NS
Number of relapses in the 2 ys before screening[a]	2.0 (1.0, 2.0)	2.0 (2.0, 3.0)	≤0.0001
Prior MS treatment, n (%)	85 (51.8)	65 (59.1)	NS
EDSS score at BL	2.0 (1.5, 3.5)	2.5 (1.5, 3.5)	≤0.05
EDSS score at M12	2.0 (1.5, 3.0)	2.3 (1.5, 3.5)	≤0.05
EDSS score at M24	2.0 (1.5, 3.0)	2.0 (1.5, 3.5)	≤0.05
PASAT score at BL	52.0 (45.0, 57.0)	52.0 (44.0, 57.0)	NS
PASAT score at M12	54.0 (48.0, 58.0)	52.0 (46.5, 58.0)	NS
PASAT score at M24	56.0 (49.0, 59.0)	54.0 (44.0, 57.0)	≤0.05
T25FWT score at BL	4.7 (4.1, 5.9)	5.1 (4.3, 6.8)	≤0.05
T25FWT score at M12	4.8 (4.2, 6.0)	4.9 (4.3, 6.4)	NS
T25FWT score at M24	4.8 (4.2, 5.6)	5.0 (4.2, 6.3)	NS
9HPT score at BL	19.7 (18.1, 22.7)	21.3 (19.0, 24.8)	≤0.05
9HPT score at M12	19.6 (17.8, 22.3)	20.7 (18.0, 24.3)	NS
9HPT score at M24	19.3 (17.7, 22.1)	20.4 (17.6, 24.0)	NS
NBV (cm^3)	1,524.4 (1,475.2, 1,572.3)	1,520.0 (1,453.9, 1,585.3)	NS
Change in brain volume from BL to M12 (%)	−0.20 (−0.60, 0.10)	−0.55 (−1.1, −0.17)	≤0.001
Change in brain volume from BL to M24 (%)	−0.44 (−1.0, −0.1)	−1.10 (−1.8, −0.5)	≤0.0001
Presence of Gd+ T1 lesions at BL, n (%)	39 (23.9)	69 (63.3)	≤0.0001
Number of Gd+ T1 lesions at BL	0 (0, 0)	1 (0, 3.0)	≤0.0001
T2LV at BL (cm^3)	1.97 (0.82, 4.86)	6.12 (2.72, 12.48)	≤0.0001
Change in T2LV from BL to M12 (cm^3)	0.01 (−0.12, 0.15)	0.01 (−0.51, 0.49)	NS
Change in T2LV from BL to M24 (cm^3)	0.02 (−0.13, 0.28)	−0.06 (−0.76, 0.42)	NS
Follow-up duration (y)	8.8 (3.6, 9.2)	8.7 (3.7, 9.3)	NS
Patients who completed M24, n (%)	149 (90.9)	95 (86.4)	NS
Patients who completed M48, n (%)	125 (76.2)	84 (76.4)	NS
Patients who completed M84, n (%)	109 (66.5)	70 (63.6)	NS
Patients who completed M96, n (%)	104 (63.4)	66 (60.0)	NS
NfL, geometric mean (pg/mL)			
BL	19.07	57.47	≤0.0001
M12	15.67	21.79	≤0.0001
M24	15.31	21.76	≤0.05
BL-M12	16.82	32.22	≤0.0001
BL-M24	16.42	27.36	≤0.0001

Abbreviations: 9HPT = 9-Hole Peg Test; BL = baseline; EDSS = Expanded Disability Status Scale; Gd+ = gadolinium enhancing; M = month; NBV = normalized brain volume; NfL = neurofilament light chain; NS = not significant; PASAT = Paced Auditory Serial Addition Test; T25FWT = Timed 25-Foot Walk Test; T2LV = T2 lesion volume; Q, quartile.
Data are presented as median (Q1, Q3), unless stated otherwise.
[a] Mean ± SD number of relapses in the 2 years before screening: 1.9 ± 0.90 in the <30 pg/mL group and 2.6 ± 1.38 in the ≥30 pg/mL group

Figure 1 Kaplan-Meier plots of time to event by NfL assessment for disability outcomes (A) at BL, (B) over 12 months, and (C) over 24 months



The reference visit is defined as BL for A, M12 for B, and M24 for C. (1.1) EDSS score ≥4 (only patients with an initial BL EDSS <4 were analyzed); (1.2) 6m-CDW (change of ≥1.5 in EDSS score if initial EDSS = 0, ≥1 if initial EDSS between 1 and 5, or ≥0.5 if initial EDSS >5); (1.3) 20% worsening in the T25FWT; (1.4) 20% worsening in the 9HPT; and (1.5) 20% worsening in the PASAT (only patients with an initial PASAT score >0 were analyzed). 6m-CDW = 6-month confirmed disability worsening; 9HPT = 9-Hole Peg Test; BL = baseline; EDSS = Expanded Disability Status Scale; M = month; NfL = neurofilament light chain; PASAT = Paced Auditory Serial Addition Test; T25FWT = Timed 25-Foot Walk Test.

−2.07 to −0.17), with the difference being statistically significant from year 3 onward (figure 2, A). Qualitatively similar, though not always significant, trends were observed when patients were stratified by the geometric mean of NfL taken over either 12 or 24 months (figure 2, B–C). It is of note that the number and proportion of patients categorized as having high NfL was higher at BL before initiation of study treatment (figure 2, A) compared with NfL assessments taken during the fingolimod treatment phase (figure 2, B–C).

Prognostic value of NfL in different predictor sets for long-term outcomes

Combination of NfL with clinical and/or MRI measures

The additional value of NfL to predict changes in brain volume and long-term clinical outcomes over conventional clinical and/or MRI measures is illustrated in table 4 and figure 3. Regardless of NfL measured at a single time point or integral measures over 12 or 24 months, the area under the ROC curve was generally lowest for models that used only clinical measures (CM; AUC range, 0.599–0.873), intermediate for models that combined clinical measures and NfL (CM + NfL; AUC range, 0.623–0.927) or clinical and MRI measures (CM + MRI; AUC range, 0.653–0.939), and highest for models that used clinical and MRI measures in combination with NfL (CM + MRI + NfL; AUC range, 0.658–0.954).

The best prognostic results for the long-term outcomes were achieved when an integral of NfL_{long} over 24 months was combined with clinical and MRI parameters, indicating that both MRI and NfL have additional value when each is combined with clinical measures, but that NfL has additional, qualitatively different prognostic value over conventional clinical and MRI measures.

Single NfL at BL vs integral measures over 12 and 24 months

An integral measure of serial NfL assessments was superior for the prognosis of long-term outcomes in MS compared with measuring NfL only once (table 4). Models that used only a single assessment of NfL at BL had a lower AUC compared with models that used an integral measure of NfL over time (BL-month 12 and BL-month 24).

The best prognostic results for the long-term outcomes were achieved when an integral measure of serial NfL was taken over 24 months in combination with clinical and MRI parameters. The area under the ROC curve for the CM + MRI + NfL model was 0.834 for ARBA, 0.954 for reaching EDSS ≥4, 0.849 for 6m-CDW, 0.868 for 20% worsening in the T25FWT, 0.777 for 20% worsening in the 9HPT, and 0.875 for 20% worsening in the PASAT.

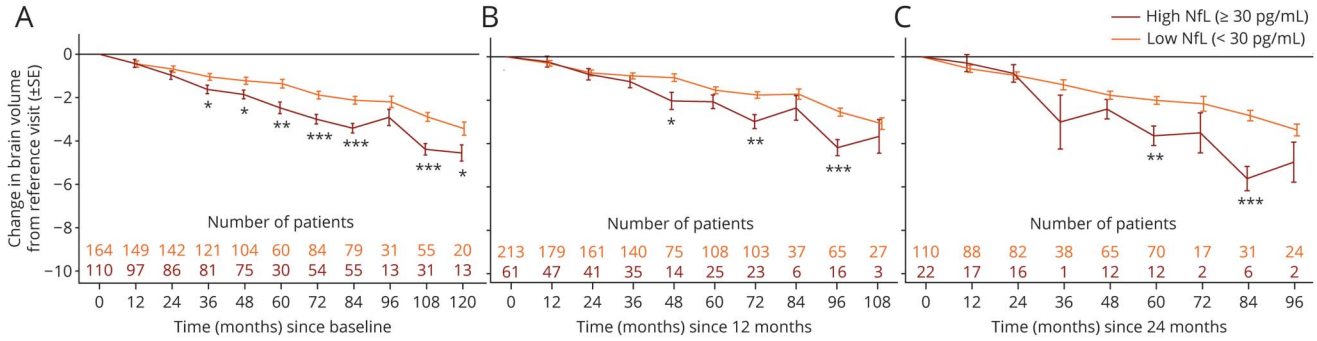
Discussion

NfL has been established as the first blood-based biomarker for MS to reflect current disease activity (relapses and lesion formation) and therapy response; moreover, NfL is able to predict—on the group level—the degree of long-term disability and features of neuronal degeneration based on BL measurements before starting disease-modifying therapies.[6,8–12,21,22] However, this approach does not factor in post-BL treatment effects for the prediction of long-term outcomes, and the accuracy of single-time NfL assessments could be limited by their short-term fluctuations due to intercurrent acute disease activity.

This analysis from the pooled fingolimod phase 3 clinical program is the first to address these issues and demonstrates that NfL_{long} over 12 or 24 months is superior to single BL NfL measures. The combination of NfL_{long} with clinical and MRI measures further improves the ability to predict the 10-year disability outcomes for patients with RRMS.

NfL reflects different disease features compared with MRI; 36.7% of patients whose brain scans were free of Gd+ lesions at BL had NfL concentrations categorized as high. Plausible

Figure 2 Estimated mean PBVC from BL by NfL assessment (A) at BL, (B) over 12 months, and (C) over 24 months



The reference visit is defined as BL for (A), M12 for (B), and M24 for (C). In (A), where the categorization was performed by BL NfL (before study drug initiation), more patients were categorized as having high NfL (n = 110) compared with (B) (n = 61) and (C) (n = 22) where patients were categorized by a geometric mean under fingolimod treatment. *$p \le 0.05$, **$p \le 0.001$, and ***$p \le 0.0001$ for high vs low NfL. BL = baseline; M = month; NfL = neurofilament light chain; PBVC = percentage brain volume change; SE = standard error.

Table 4 Area under the ROC curve for different predictor sets of clinical, NfL, and MRI parameters for the prognosis of long-term clinical outcomes and brain volume change at M84

Outcomes (predictor sets)	BL	Geometric mean over 12 mo	Geometric mean over 24 mo
EDSS score ≥4			
CM	0.840	0.873	0.849
CM + NfL	0.874	0.877	0.927
CM + MRI	0.842	0.939	0.867
CM + MRI + NfL	0.882	0.945	0.954
6m-CDW			
CM	0.699	0.631	0.681
CM + NfL	0.715	0.642	0.781
CM + MRI	0.718	0.678	0.756
CM + MRI + NfL	0.739	0.683	0.849
20% worsening in the T25FWT			
CM	0.652	0.599	0.617
CM + NfL	0.659	0.623	0.778
CM + MRI	0.653	0.697	0.686
CM + MRI + NfL	0.658	0.720	0.868
20% worsening in the 9HPT			
CM	0.652	0.603	0.605
CM + NfL	0.691	0.682	0.702
CM + MRI	0.694	0.674	0.618
CM + MRI + NfL	0.746	0.745	0.777
20% worsening in the PASAT			
CM	0.644	0.641	0.702
CM + NfL	0.667	0.635	0.740
CM + MRI	0.704	0.733	0.789
CM + MRI + NfL	0.715	0.758	0.875
ARBA[a] ≤ −0.4			
CM	0.737	0.709	0.711
CM + NfL	0.760	0.733	0.761
CM + MRI	0.743	0.734	0.790
CM + MRI + NfL	0.762	0.781	0.834

Abbreviations: 6m-CDW = 6-month confirmed disability worsening; 9HPT = 9-Hole Peg Test; ARBA = annualized rate of brain atrophy; BL = baseline; CM = clinical model; EDSS = Expanded Disability Status Scale; NfL = neurofilament light chain; PASAT = Paced Auditory Serial Addition Test; PBVC, percentage brain volume change; ROC = receiver operating characteristic; SIENA = structural image evaluation using normalization of atrophy; T25FWT = Timed 25-Foot Walk Test.

[a] SIENA PBVC was converted to ARBA by ([SIENA/100 + 1] [365.25/days] −1) × 100, where "SIENA" represents the PBVC obtained between 2 scans and "days" means the days for the scan date relative to day 1,[32] categorized as ≤−0.4% vs >0.4% to correct for differences in the distance between MRI between patients.

causes for this constellation include lesion formation in the spinal cord, disease activity that escapes detection on routine MRI,[21] brain diffuse pathology in the gray and/or white matter, or early phases of lesion formation not yet visible in routine MRI.[23] More research with high-frequency MRI and NfL measurements is needed to better understand the kinetics of change in NfL concentration in blood in relation to MS lesion formation in the CNS.

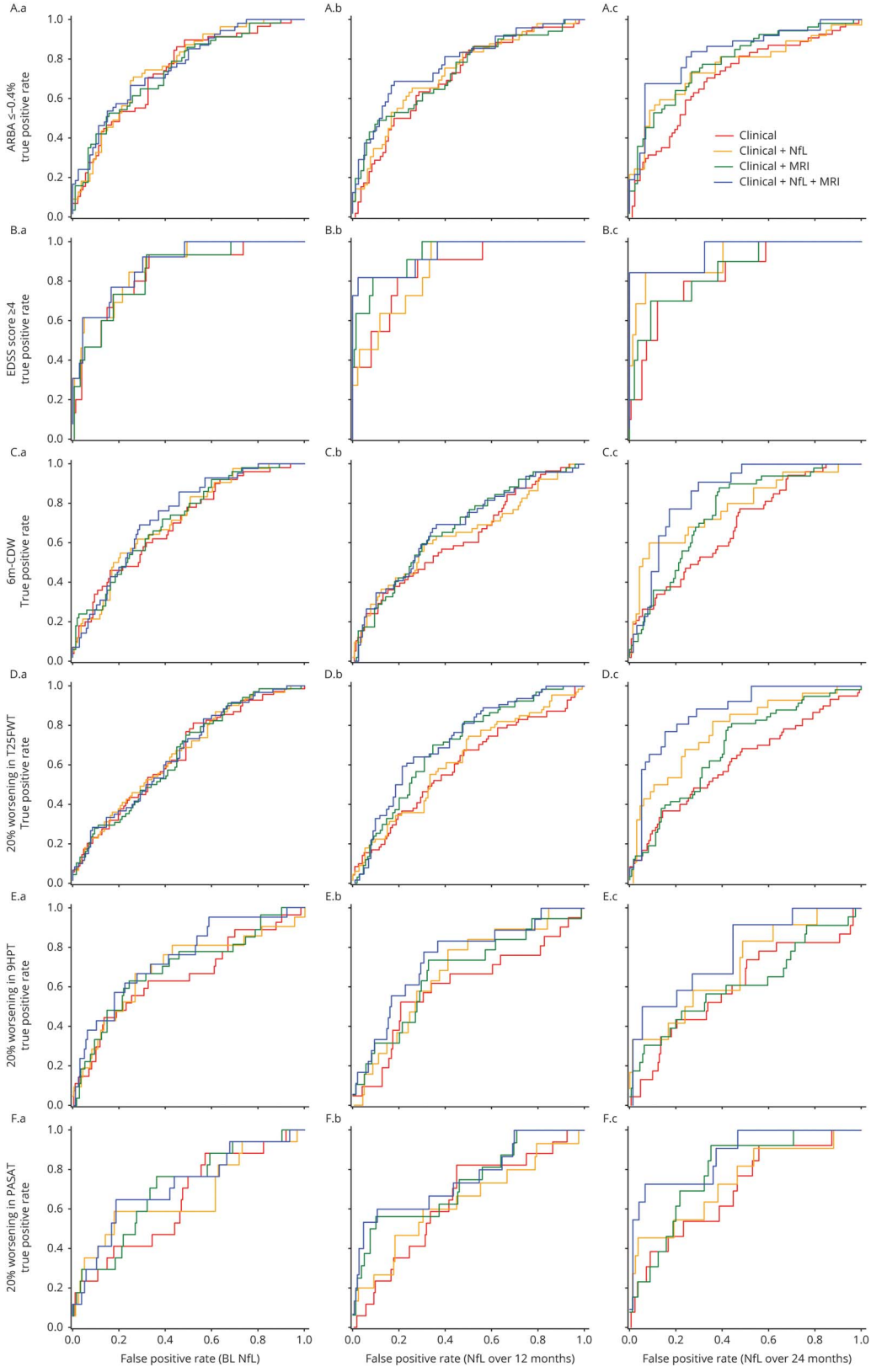
The BL features of all patients who contributed to this NfL analysis were not notably different from the overall population of FREEDOMS and TRANSFORMS and representative of a typical RRMS population. However, patients with high NfL at BL represented a more active and advanced MS population.

Patients with high NfL at BL had a higher on-study PBVC for up to 120 months. This is clinically relevant because NBV has been shown to predict long-term outcomes in MS.[24–27] The prognostic value of BL NfL for on-study BVL observed in the current study was broadly in line with recently published work, partly using the same data from FREEEDOMS and TRANSFORMS.[10] Furthermore, the present results are largely consistent with the Comprehensive Longitudinal Investigations in MS at Brigham and Women's Hospital (CLIMB)[22] and Expression, Proteomics, Imaging, Clinical (EPIC)[28] studies. The CLIMB study reported a correlation of early annual and averaged yearly serum NfL levels with 10-year MRI outcomes and worsening of fatigue measures.[22] In the EPIC study, BL serum NfL levels were predictive of brain atrophy in the following 2–10 years.[28] Of interest, we identified a lag in time between NfL and BVL in our study, suggesting that these measures differ in their kinetic change. Although the curves of BVL separated almost immediately when categorizing patients by their BL NfL values, a longer lag time was identified when categorizing patients by NfL_{long} assessment over 12 or 24 months. Acute MS disease activity (e.g., Gd+ lesions) could be one explanation for high NfL values in a single NfL assessment at BL, and Gd+ lesions have been identified as a strong predictor for on-study BVL in 3 phase 3 trials.[29]

Consistently, the geometric mean of NfL_{long} was found to be superior for the prognosis of unfavorable disability outcomes compared with single NfL measures at BL. The prediction of the long-term outcomes based on elevated NfL_{long} is less influenced by an intermittent increase of disease activity and hence may better reflect the chronic process of neuronal injury and eventual tissue loss.

The area under the ROC curve analysis demonstrated that long-term outcomes were better predicted when MRI and clinical features were combined with NfL_{long} compared with when the former 2 were used alone, indicating that NfL_{long} identifies an additional pathogenesis that escapes the current standard. The conceptual advantage of NfL_{long} over single NfL measures at BL is the inclusion of the disease-modifying effect of therapies as an additional factor defining the long-term outcomes. Based on these findings, NfL has been

Figure 3 ROC curves for analyses of ARBA, EDSS ≥4, 6m-CDW, and 20% worsening in the T25FWT, 9HPT, and PASAT at M84



Models with/without predictor NfL (A) at BL, (B) over 12 months, and (C) over 24 months, for (3.1) ARBA up to −0.4%, (3.2) EDSS score ≥4, (3.3) 6m-CDW, (3.4) 20% worsening in the T25FWT, (3.5) 20% worsening in the 9HPT, and (3.6) 20% worsening in the PASAT. 6m-CDW = 6-month confirmed disability worsening; 9HPT = 9-Hole Peg Test; ARBA = annualized rate of brain atrophy; BL = baseline; EDSS = Expanded Disability Status Scale; M = month; NfL = neurofilament light chain; PASAT = Paced Auditory Serial Addition Test; ROC = receiver operating characteristic; T25FWT = Timed 25-Foot Walk Test.

suggested as an end point for phase 2 studies in progressive MS,[30] where we currently lack established trial paradigms.[31]

The sample size of this post hoc analysis was limited by the availability of blood samples, and the current study was not powered to show the effects on long-term outcomes in some of the disability measures. Despite these limitations, a consistent trend toward unfavorable long-term outcomes in patients with high NfL with HRs up to a factor of 8 between patients with high compared with low NfL was found, suggesting that NfL is a promising biomarker to stratify patients into risk groups. Given the limited sample size, the focus on only 1 disease-modifying therapy (fingolimod), the post hoc nature of this study, and disease heterogeneity, confirmatory evidence for the value of NfL for the long-term prognosis of patients with MS is needed from future prospective clinical studies with long-term data collection.

An integral measure of longitudinal NfL assessments collected over 12 or 24 months might improve the accuracy of the prognosis of long-term disability outcomes in patients with MS. In the current study, the highest prognostic value was achieved when an integral measure of NfL in combination with clinical and MRI features was used. The prognostic value of low NfL concentrations for beneficial long-term outcomes on the group level also supports the need to keep NfL levels low in individual patients. Thus, NfL in blood fulfills a critical requirement as a prognostic biomarker for disability worsening and could be useful in monitoring treatment success in patients with MS.

Study funding

The study was funded by Novartis Pharma AG. The study sponsor participated in the design and conduct of the study, data collection, data management, data analysis and interpretation, and preparation, review, and approval of the manuscript. The biostatistical analyses were performed at DATAMAP GmbH, Freiburg, Germany.

Disclosure

D.A. Häring and H. Kropshofer are employees of Novartis Pharma AG. L. Kappos' institution (University Hospital Basel) has received the following in the last 3 years and used exclusively for research support: steering committee, advisory board, and consultancy fees from Actelion, Addex, Bayer HealthCare, Biogen Idec, Biotica, Genzyme, Lilly, Merck, Mitsubishi, Novartis, Ono Pharma, Pfizer, Receptos, Sanofi, Santhera, Siemens, Teva, UCB, and XenoPort; speaker fees from Bayer HealthCare, Biogen Idec, Merck, Novartis, Sanofi, and Teva; support for educational activities from Bayer HealthCare, Biogen, CSL Behring Genzyme, Merck, Novartis, Sanofi, and Teva; royalties from Neurostatus products; license fees for Neurostatus products; and grants from Bayer HealthCare, Biogen Idec, European Union, INNO-Swiss, Merck, Novartis, Roche Research Foundation, Swiss MS Society, and the Swiss National Research Foundation. J.A. Cohen has received personal compensation for consulting for Convelo and Population Council; speaking for Mylan; and serving as an Editor of *Multiple Sclerosis Journal*. A. Shah is an employee of Novartis Healthcare Pvt. Ltd. R. Meinert is an employee of DATAMAP GmbH, which provides services to Novartis Pharma AG. D. Leppert was an employee of Novartis Pharma AG at the time of outline development (until January 2019); he has received personal compensation for consulting and speaking and travel reimbursement from Quanterix, Orion, and Sanofi. D. Tomic was an employee of Novartis Pharma AG at the time of submission of the manuscript. J. Kuhle's institution (University Hospital Basel) received and used exclusively for research support: consulting fees from Biogen, Novartis, Roche, and Teva; speaker fees from the Swiss MS Society, Biogen, Novartis, Roche, and Sanofi; travel expenses from Merck, Novartis, and Roche; and grants from ECTRIMS Research Fellowship Programme, University of Basel, Swiss MS Society, Swiss National Research Foundation (320030_189140/1), Bayer, Biogen, Celgene, Genzyme, Merck, Novartis, Roche, and Sanofi.Go to Neurology.org/NN for full disclosures.

Publication history

Received by *Neurology: Neuroimmunology & Neuroinflammation* March 30, 2020. Accepted in final form June 19, 2020.

Appendix Authors

Name	Location	Contribution
Dieter A. Häring, PhD	Novartis Pharma AG, Basel, Switzerland	Study concept, design, execution, data acquisition, analysis and interpretation, outline review, critical revision of the manuscript, and statistical analysis
Harald Kropshofer, PhD	Novartis Pharma AG, Basel, Switzerland	Study concept, design, execution, data acquisition, analysis and interpretation, critical revision of the manuscript, obtaining study funding, and supervising the research
Ludwig Kappos, MD	Research Center for Clinical Neuroimmunology and Neuroscience Basel and Departments of Medicine, Clinical Research, Biomedicine and Biomedical Engineering, University Hospital and University of Basel, Switzerland	Study concept, design, data analysis and interpretation, and critical revision of the manuscript
Jeffrey A. Cohen, MD	Department of Neurology, Mellen MS Center, Neurological Institute, Cleveland Clinic, OH	Study concept, design, execution, data acquisition, analysis and interpretation, outline review, and critical revision of the manuscript
Anuja Shah, PhD	Novartis Healthcare Pvt. Ltd. Hyderabad, India	Conducted literature search, manuscript drafting, revising, and editing

Continued

Appendix *(continued)*

Name	Location	Contribution
Rolf Meinert, PhD	DATAMAP GmbH, Freiburg, Germany	Data analysis and interpretation, outline review, critical revision of the manuscript, and statistical analysis
David Leppert, MD	Research Center for Clinical Neuroimmunology and Neuroscience Basel and Departments of Medicine, Clinical Research, Biomedicine and Biomedical Engineering, University Hospital and University of Basel, Switzerland	Study concept, execution, data acquisition and interpretation, critical revision of the manuscript, obtaining study funding, technical support, and supervising the research
Davorka Tomic, DVM, PhD	Novartis Pharma AG, Basel, Switzerland	Study concept, design, execution, data analysis and interpretation, outline review, critical revision of the manuscript, obtaining study funding, technical support, and supervising the research
Jens Kuhle, MD, PhD	Neurologic Clinic and Policlinic, Departments of Medicine, Biomedicine and Clinical Research, University Hospital and University of Basel, Switzerland	Study concept, data acquisition and interpretation, outline review, critical revision of the manuscript, obtaining study funding, technical support, and supervising the research

References

1. Antel J, Antel S, Caramanos Z, Arnold DL, Kuhlmann T. Primary progressive multiple sclerosis: part of the MS disease spectrum or separate disease entity? Acta Neuropathol 2012;123:627–638.
2. Lassmann H, van Horssen J, Mahad D. Progressive multiple sclerosis: pathology and pathogenesis. Nat Rev Neurol 2012;8:647–656.
3. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 2014;83:278–286.
4. Yan Y, Jensen K, Brown A. The polypeptide composition of moving and stationary neurofilaments in cultured sympathetic neurons. Cell Motil Cytoskeleton 2007;64:299–309.
5. Teunissen CE, Khalil M. Neurofilaments as biomarkers in multiple sclerosis. Mult Scler 2012;18:552–556.
6. Khalil M, Teunissen CE, Otto M, et al. Neurofilaments as biomarkers in neurological disorders. Nat Rev Neurol 2018;14:577–589.
7. Kuhle J, Barro C, Disanto G, et al. Serum neurofilament light chain in early relapsing remitting MS is increased and correlates with CSF levels and with MRI measures of disease severity. Mult Scler 2016;22:1550–1559.
8. 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.
9. 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.
10. 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.
11. Novakova L, Zetterberg H, Sundstrom P, et al. Monitoring disease activity in multiple sclerosis using serum neurofilament light protein. Neurology 2017;89:2230–2237.
12. Piehl F, Kockum I, Khademi M, et al. Plasma neurofilament light chain levels in patients with MS switching from injectable therapies to fingolimod. Mult Scler 2018;24:1046–1054.
13. Siller N, Kuhle J, Muthuraman M, et al. Serum neurofilament light chain is a biomarker of acute and chronic neuronal damage in early multiple sclerosis. Mult Scler 2019;25:678–686.
14. Kappos L, Radue EW, O'Connor P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. N Engl J Med 2010;362:387–401.
15. Cohen JA, Barkhof F, Comi G, et al. Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. New Engl J Med 2010;362:402–415.
16. Kappos L, O'Connor P, Radue EW, et al. Long-term effects of fingolimod in multiple sclerosis: the randomized FREEDOMS extension trial. Neurology 2015;84:1582–1591.
17. Cohen JA, Khatri B, Barkhof F, et al. Long-term (up to 4.5 years) treatment with fingolimod in multiple sclerosis: results from the extension of the randomised TRANSFORMS study. J Neurol Neurosurg Psychiatry 2016;87:468–475.
18. Cohen JA, Tenenbaum N, Bhatt A, Zhang Y, Kappos L. Extended treatment with fingolimod for relapsing multiple sclerosis: the 14-year LONGTERMS study results. Ther Adv Neurol Disord 2019;12:1756286419878324.
19. Kappos L, Mehling M, Arroyo R, et al. Randomized trial of vaccination in fingolimod-treated patients with multiple sclerosis. Neurology 2015;84:872–879.
20. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J Radiol 2004;5:11–18.
21. Ruggieri SLA, Tinelli E, Giglio De L, Prosperini L, Gasperini C, Pozzilli C. Measuring disease activity in multiple sclerosis: the essential role of spinal cord MRI monitoring. Mult Scler J 2018;24:121–327.
22. Chitnis T, Gonzalez C, Healy BC, et al. Neurofilament light chain serum levels correlate with 10-year MRI outcomes in multiple sclerosis. Ann Clin Transl Neurol 2018;5:1478–1491.
23. Werring DJ, Brassat D, Droogan AG, et al. The pathogenesis of lesions and normal-appearing white matter changes in multiple sclerosis: a serial diffusion MRI study. Brain 2000;123(pt 8):1667–1676.
24. Sormani MP, Kappos L, Radue EW, et al. Defining brain volume cutoffs to identify clinically relevant atrophy in RRMS. Mult Scler 2017;23:656–664.
25. Traboulsee AL, Cornelisse feminine P, Sandberg-Wollheim M, et al. Prognostic factors for long-term outcomes in relapsing-remitting multiple sclerosis. Mult Scler J Exp Transl Clin 2016;2:2055217316666406.
26. Miller DH, Lublin FD, Sormani MP, et al. Brain atrophy and disability worsening in primary progressive multiple sclerosis: insights from the INFORMS study. Ann Clin Transl Neurol 2018;5:346–356.
27. Gaetano L, Haring DA, Radue EW, et al. Fingolimod effect on gray matter, thalamus, and white matter in patients with multiple sclerosis. Neurology 2018;90:e1324–e1332.
28. Canto 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 2019;76:1359–1366.
29. Radue EW, Barkhof F, Kappos L, et al. Correlation between brain volume loss and clinical and MRI outcomes in multiple sclerosis. Neurology 2015;84:784–793.
30. Sormani MP, Haering DA, Kropshofer H, et al. Blood neurofilament light as a potential endpoint in phase 2 studies in MS. Ann Clin Translational Neurol 2019;92:e1007–e1015.
31. Leppert D, Kuhle J. Blood neurofilament light chain at the doorstep of clinical application. Neurol Neuroimmunol Neuroinflamm 2019;6:e599. doi: 10.1212/NXI.0000000000000599.
32. Jeffery DR, Di Cantogno EV, Ritter S, Meier DP, Radue EW, Camu W. The relationship between the rate of brain volume loss during first 24 months and disability progression over 24 and 48 months in relapsing MS. J Neurol 2016;263:299–305.

Long-term prognostic value of longitudinal measurements of blood neurofilament levels

Dieter A. Häring, Harald Kropshofer, Ludwig Kappos, et al.

Neurol Neuroimmunol Neuroinflamm 2020;7;
DOI 10.1212/NXI.0000000000000856