RESEARCH ARTICLE OPEN ACCESS

Clinical Utility of Serum Neurofilament Light Chain and Glial Fibrillary Acidic Protein for Tracking Disease Activity in Multiple Sclerosis

Maria Martinez-Serrat,[1,2] Cansu Tafrali,[1,2] Rina Demjaha,[1,2] Timothy Kaiser,[1,2] Michaela Tanja Haindl,[1,2] Birgit Helmlinger,[1,3] Edith Hofer,[1,4] Stefan Ropele,[1] Bettina Heschl,[1] Anna Damulina,[1,2] Emilia Platos,[1] Daniela Pinter,[1,3] Felix Wohlrab,[5,6] Tatiana Usnich,[5] Friedemann Paul,[5,6,7] David Leppert,[8,9] Pascal Benkert,[8,9] Jens Kuhle,[8,9] Christian Enzinger,[1] and Michael Khalil[1,2]

Correspondence
Dr. Khalil
michael.khalil@medunigraz.at

Neurol Open Access 2026;2:e000161. doi:10.1212/WN9.0000000000000161

Abstract

Background and Objectives

Monitoring disease activity in patients with multiple sclerosis (MS) remains challenging, and studying blood-based biomarkers that improve prognostic accuracy in routine clinical practice is needed. The objective of this study was to investigate the dynamic behavior of serum neurofilament light chain (sNfL) and serum glial fibrillary acidic protein (sGFAP) throughout the timeline of clinical relapses. In addition, we assessed which sampling time points provide the most relevant prognostic information to identify patients at risk of future disease activity. This study also assessed the influence of high-efficacy therapies.

Methods

We conducted a longitudinal cohort study using serum samples from patients followed at the MS outpatient clinic of the Medical University of Graz, Austria. Inclusion criteria were age older than 18 years, diagnosis of MS according to the 2024 McDonald criteria, and availability of at least 5 serum samples over a minimum of 5 years. sNfL and sGFAP were measured from the repeated sampling using Simoa HD-X and correcting for age, sex, and BMI with Z-scores calculated from a large healthy aging reference cohort.

Results

The study included 160 patients with MS (1,228 samples, 64.4% female; mean age 32.5 years; median follow-up of 10.4 years). sNfL Z-scores showed a nonlinear increase before ($p < 0.05$), during, and up to 1 year after a relapse (all $p < 0.001$). sGFAP Z-scores were increased during a relapse and up to 2 years afterward (all $p < 0.01$), although with less consistent temporal patterns. When measuring sNfL 9 months after a relapse, higher values (Z-score >1.5) were associated with the highest odds ratio ([OR] 2.94 [95% CI 1.73–5.0]; $p < 0.001$) of developing future disease activity within the next 1.5 years. In this context, sGFAP did not show any predictive value.

Discussion

These findings provide a better insight into the dynamics of these proteins, highlighting that sNfL provides clinically meaningful short-term prognostic information when measured during remission, whereas sGFAP shows limited prognostic utility in this context, despite detectable changes around relapses. Our results support the integration of optimized sampling strategies into routine MS monitoring to improve individualized care.

[1]Department of Neurology, Medical University of Graz, Austria; [2]Neurology Biomarker Research Unit, Medical University of Graz, Austria; [3]Research Unit for Neuronal Plasticity and Repair, Medical University of Graz, Austria; [4]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria; [5]Experimental and Clinical Research Center, A Cooperation Between Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité–Universitätsmedizin Berlin, Germany; [6]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Experimental and Clinical Research Center, Berlin, Germany; [7]NCRC-Neuroscience Clinical Research Center, Charite–Universitätsmedizin Berlin, Corporate Member of Freie Universitat Berlin, Humboldt-Universitat zu Berlin, Germany; [8]Multiple Sclerosis Centre and Research Center for Clinical Neuroimmunology and Neuroscience (RC2NB), Departments of Biomedicine and Clinical Research, University Hospital and University of Basel, Switzerland; and [9]Department of Neurology, University Hospital and University of Basel, Switzerland.

The Article Processing Charge was funded by the authors.

Glossary

AIC = Akaike Information Criterion; **BIC** = Bayesian Information Criterion; **CV** = coefficient of variation; **EDA-3** = evidence of disease activity-3; **EDSS** = Expanded Disability Status Scale; **GAM** = generalized additive mixed model; **HE-DMT** = high-efficacy disease-modifying therapy; **IQR** = Interquartile range; **LMM** = linear mixed-effects model; **MS** = multiple sclerosis; **NEDA-3** = no evidence of disease activity-3; **OR** = odds ratio; **sGFAP** = serum glial fibrillary acidic protein; **sNfL** = serum neurofilament light chain.

Introduction

Multiple sclerosis (MS) is an immune-mediated disease affecting the CNS and represents the leading cause of neurologic disability in young adults.[1] Monitoring disease activity in routine clinical practice remains challenging because of its complex and dynamic pathophysiology. Detecting future disease activity on time is critical for guiding clinical decisions and preventing irreversible disability. Although advances in imaging and clinical assessments have improved disease monitoring, they often lack the sensitivity required for routine clinical practice. Recent efforts have therefore focused on identifying complementary blood-based biomarkers capable of capturing both neuroinflammatory and neurodegenerative processes.[2-4] Among the most promising are serum neurofilament light chain (sNfL) and serum glial fibrillary acidic protein (sGFAP). Establishing robust interpretative guidelines for these biomarkers is essential to support individualized disease management and therapeutic decision making, especially as their increasing validation now places them on the doorstep of clinical implementation. Clear and accurate guidance for interpreting sNfL and sGFAP in clinical settings is therefore critically needed.[2-4]

sNfL, a marker of axonal injury, is strongly associated with clinical relapses and MRI activity in MS.[5-7] Elevated CSF and serum levels are consistently observed during relapses compared with remission or healthy controls.[8-10] More recently, sNfL has emerged as a promising prognostic marker, with evidence supporting its ability to predict relapses and other aspects of disease activity within the following year.[11,12] Several studies emphasized the importance of frequent sNfL monitoring to complement clinical and radiologic assessments.[13-16] Repetitive sampling has shown that sNfL can rise months before a relapse, peaking around clinical onset, whereas other studies demonstrate that sNfL elevations closely track gadolinium-enhancing (Gd+) lesions within a 3-month window.[13-16] In line with these observations, recent guidelines recommend repetitive measurements every 3–6 months.[17] However, some studies rely on small cohorts or specific settings. The short-term dynamics of sNfL fluctuations in relation to clinical relapses and their optimal sampling intervals for prognostic purposes remain poorly understood.

sGFAP, an astrocytic marker, reflects astrocytic activation and gliosis and has been associated with progression independent of relapses, a major contributor to long-term disability in MS.[18-21] Evidence suggests that sGFAP may have prognostic value for disease progression,[22] particularly in patients with low sNfL levels,[21] indicating complementary pathologic insights. Evidence regarding the role of GFAP in clinical relapses is inconsistent: Ahile a couple of studies reported increased CSF or serum GFAP during relapses compared with remission,[8,22] others found no elevation of sGFAP before relapse.[12,21] Although sGFAP seems relevant for long-term progression, its short-term fluctuations and potential association with future disease activity remain unclear.

In this study, we focus on longitudinal sNfL and sGFAP measurements to investigate their dynamic changes during relapsing activity in patients with MS. Our objectives are to characterize patterns associated with relapse activity, assess optimal sampling windows, and evaluate the potential of these proteins to predict future disease activity as assessed by the evidence of disease activity-3 (EDA-3) criteria. In addition, given the strong capacity of high-efficacy disease-modifying therapy (HE-DMT) to suppress the acute focal inflammation, we examine their impact on the prognostic value of these biomarkers.

Methods

Study Design and Participants

This longitudinal study included 160 patients with MS recruited from the routine MS outpatient clinic of the Department of Neurology, Medical University of Graz, Austria. All data from patients followed at the MS outpatient clinic were routinely stored in a prospective single-center database. For the present analysis, patients were identified through this database based on the availability of longitudinal serum samples meeting the predefined inclusion criteria. The inclusion criteria for the participants in this cohort were age older than 18 years at baseline, diagnosed with MS according to the 2024 McDonald criteria, and availability of at least 5 serum samples over a minimum period of 5 years. After inclusion, 1 patient with primary progressive MS was identified and excluded from the analyses.

The cohort comprises patients with different MS disease courses, including relapsing MS and secondary progressive MS.[23]

Demographic and clinical data included age, sex, relapse dates and frequency, Expanded Disability Status Scale (EDSS) scores[24] in remission, disease duration, and DMT.

Sampling and Serum Measurements

Serum samples were obtained through peripheral blood collection following international consensus guidelines.[25] sNfL

and sGFAP concentrations were quantified in duplicate using a 2-plex ultrasensitive single molecule array assay (Simoa N2PB Advantage PLUS, Quanterix; product number: 104670) on an HD-X platform, according to manufacturer's instructions. All duplicates included in the analysis had a coefficient of variation (CV) < 20%. Each run included 2 internal quality control samples, with an interassay CV <20%.

Z-scores were calculated to adjust for confounding factors (age and sex for sNfL and age, sex, and body mass index for sGFAP) using a large healthy aging reference cohort.[11,26,27] Z-scores represent the number of SDs from the reference population mean.

MRI

MRI scans were performed on a 3-Tesla Prisma fit system (Siemens Medical Systems), as described in previous work from this group.[28] The acquisition protocol included T1-weighted and T2-weighted sequences. Counts for new or enlarging T2 and T1 Gd + lesions were extracted from MRI reports by trained neurologists where new or enlarging lesions were defined relative to the previous available scan; no rebaseline imaging protocol was applied.

Clinical Evaluation

The outcome score EDA-3 was defined as the occurrence of any of the following: clinical relapses, confirmed disability worsening as assessed by EDSS scores, or new/enlarging T2 lesions or Gd + lesions. When none of these criteria were met, that was classified as having no evidence of disease activity-3 (NEDA-3).[29]

Confirmed disability worsening was defined as an increase in EDSS of ≥1.5 points from a baseline score of 0.0, ≥1.0 points from a baseline score of 1.0–5.0, or ≥0.5 point from a baseline score ≥5.5, confirmed at a subsequent visit at least 6 months later.[29] EDSS assessments were performed during scheduled clinical visits.

A relapse was defined as the onset or reoccurrence of at least 1 neurologic symptom after a stable or improving state of at least 30 days, persisting for at least 24 hours and attributed to MS. All relapses were confirmed by neurologists.[29] All patients included in the study had a relapsing disease course at some point during their clinical history. Information on relapses was obtained retrospectively from clinical records to determine relapse timing in relation to serum sampling. Not all patients experienced a clinical relapse during the period of serum collection; however, relapses occurring before that were also considered. Accordingly, all serum samples could be classified according to their time in relation to a relapse.

Statistical Analysis

Analyses were performed using IBM SPSS Statistics (version 29.0) and R (version 4.4.1); all graphs were plotted using R package ggplot2. Initially, we applied generalized additive mixed models (GAMs) for location, scale, and shape using the mgcv package in R to visualize and characterize the trajectories of sNfL and sGFAP Z-scores across the relapse timeline. This exploratory step identified the overall shape of biomarker dynamics and informed the grouping of time points into clinically meaningful intervals.

For these relapse-anchored analyses, each serum sample was assigned to a time interval based on its temporal distance from the most recent clinical relapse in the patient's clinical history. Based on these intervals, we compared the biomarker levels between them using linear mixed-effects models (LMMs) with a random intercept for patient identification to account for repeated measurements. We analyzed the differences against a predefined reference group consisting of samples collected more than 3 years after a relapse, with Bonferroni correction applied for multiple testing.

To assess the prognostic potential of both proteins, we used LMMs to identify group differences between patients showing future EDA-3 and NEDA-3. Different models were analyzed for each protein, corresponding to different definitions of remission/relapse sampling windows (3, 6, 9, 12, 18, and 24 months postrelapse), plus an overall model including all samples without relapse/remission distinction. Each model included fixed effects for the interaction of (1) time to clinical evaluation (<1 year, 1–1.5 years, 1.5–3 years, and >3 years), (2) clinical outcome at the next evaluation (EDA-3 vs NEDA-3), and (3) sampling context (relapse vs remission). Patient identification was included as a random effect with a first-order autoregressive covariance structure to account for within-subject correlation, and multiple comparisons were corrected using Bonferroni. Model fit was evaluated using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Model diagnostics were performed, and assumptions of normality and homoscedasticity were met.

To assess prognostic value, generalized linear mixed models with a binomial distribution and logit link were applied to remission samples collected within selected windows identified as significant in the LMM analysis. The outcome variable was disease activity at the next evaluation (EDA-3 vs NEDA-3), and the main predictor was protein level modeled as either a continuous variable or categorized into high vs low using predefined cutoffs. Odds ratios (ORs) with 95% CIs were reported.

Standard Protocol Approval, Registrations, and Patient Consents

The study received approval by the ethics committee of the Medical University of Graz, Austria (Ethical Approval No. 17-046 ex 05/06, 31-432 ex 18/19). All participants gave written informed consent before inclusion in the study. The study was registered at ClinicalTrials.gov under MarkMS, NCT04892134.

Data Availability

Anonymized data underlying the findings of this study are available from the corresponding author on reasonable

Table Study Cohort

	Total (N = 160)
Number of samples per patient, median (IQR)	7 (6–9)
Time between samples, y, median (IQR)	1.1 (1.0–1.9)
Follow-up time, y, median (IQR)	10.4 (7.8–13.8)
Number (%) of female	103 (64.4)
Disease course, number	
RMS	141
PMS	19
Baseline	
Age, y, mean (SD)	32.5 (7.8)
Disease duration, y, median (IQR)	2.1 (1.0–7.1)
EDSS, median (IQR)	1.0 (0.0–2.0)
Disease modifying treatment, number (%)	
No	65 (40.6)
Moderate efficacy	84 (52.5)
High efficacy	11 (6.9)
Last follow-up	
EDSS, median (IQR)	1.0 (0.0–2.5)
Disease modifying treatment, number (%)	
No	42 (26.3)
Moderate efficacy	61 (38.1)
High efficacy	57 (35.6)

Abbreviations: EDSS = Expanded Disability Status Scale; IQR = interquartile range; PMS = progressive MS; RMS = relapsing MS.

request by qualified researchers, subject to approval by the data-clearing committee of the Medical University of Graz.

Results

Baseline Characteristics

The MS outpatient clinic follows routinely over 500 patients. At the time of analysis, the number of patients with available serum samples in this database included 266 patients. Of these, 161 met the eligibility criteria regarding longitudinal serum sampling and follow-up duration. One patient with primary progressive MS was excluded, resulting in a final analytical cohort of 160 patients.

The median number of samples per patient was 7 (interquartile range [IQR]: 6–9), with a median follow-up time of 10.4 years (IQR: 7.8–13.8). The median time between samples was 1.1 years (IQR: 1.0–1.9). The cohort was 64.4% female (N = 103), with a mean age at baseline of 32.5 years (SD: 7.8). At baseline, the median disease duration was 2.1 years (IQR: 1.0–7.1) and the median EDSS score was 1.0 (IQR: 0.0–2.0); 40.6% of patients were untreated, 52.5% were taking moderate-efficacy therapies, and 6.9% were taking HE-DMTs. The disease course distribution included 141 patients with relapsing MS and 19 with progressive MS (Table).

Temporal Dynamics of Relapses

To explore how sNfL Z-score levels change around clinical relapses, we used GAMs to model their trajectory using only samples collected within 3 years before or after a relapse. The analysis showed a significant nonlinear pattern (effective degrees of freedom (edf) = 3.9, $F = 34.3$, $p < 0.001$), indicating that sNfL Z-scores do not follow a simple trend but exhibit a more complex temporal pattern (Figure 1A). As expected, the variability between individuals was also highly significant (edf = 114.8, $F = 3.7$, $p < 0.001$), reflecting substantial interpatient variability.

To further explore the clinical relevance of these changes, we grouped samples into meaningful time intervals and compared them using LMMs. Samples taken more than 3 years after a relapse served as the reference group, representing a stable state with no recent relapse activity. Compared with this reference group, sNfL Z-scores were significantly higher 6 months before a relapse ($p = 0.021$), at the time of relapse ($p < 0.001$), and up to 1 year afterward ($p < 0.001$; Figure 1C).

For sGFAP Z-scores, we also observed a statistically significant association with relapse timing (edf = 4.8, $F = 2.8$, $p = 0.013$), although the overall pattern was less consistent than for sNfL (Figure 1B). The nonlinear shape of the curve appeared less robust and may partly reflect model sensitivity rather than a true biological trajectory. As with sNfL, we observed substantial variability between patients (edf = 137.8, $F = 12.7$, $p < 0.001$).

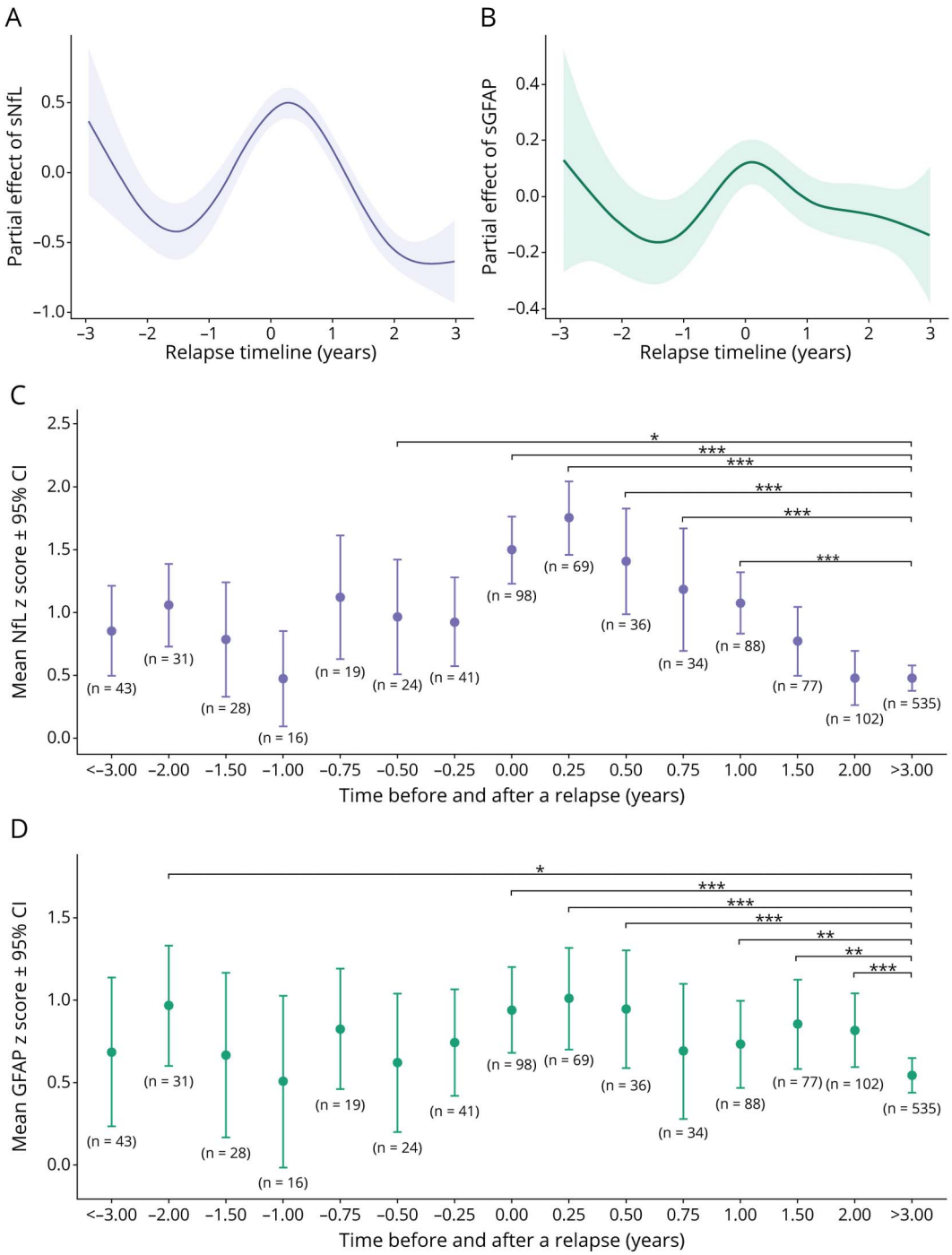
When analyzing sGFAP Z-scores across the defined time intervals, we found significant increases 2 years before a relapse ($p = 0.049$), at the time of the relapse ($p < 0.001$), and up to 2 years afterward ($p < 0.006$, Figure 1D) compared with samples taken more than 3 years after a relapse.

Including DMT classification (untreated, moderate efficacy, and high efficacy) as an additional covariate did not change the biomarker trajectories or significance. In the LMM, DMT showed a significant main effect for sNfL but no interaction with relapse-aligned time groups, and neither effect was significant for sGFAP. These findings indicate that treatment category did not meaningfully modify the temporal dynamics of either biomarker around relapses (data not shown).

Prognostic Value of Protein Z-Score Levels

To assess the clinical relevance of sNfL and sGFAP Z-scores, we first identified how the timing of biomarker measurement influences its ability to differentiate between patients who remained stable and those who later developed EDA-3 at

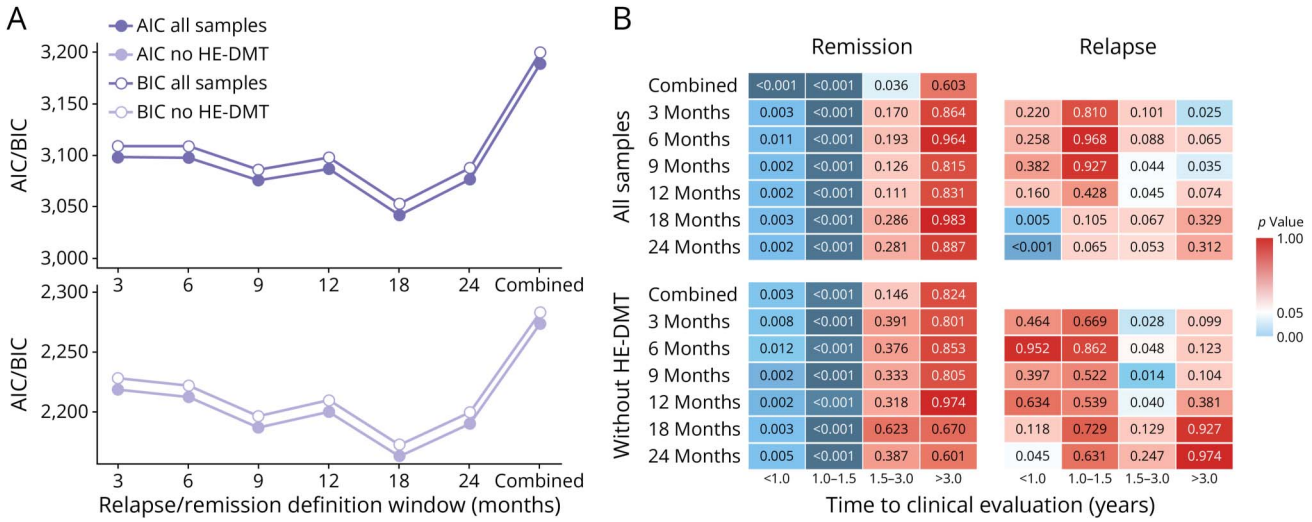
Figure 1 Temporal Dynamics of sNfL (A and C) and sGFAP (B and D) Z-Scores Around a Relapse



(A and C) Partial effects (line) and 95% CI (shade) from GAM of sNfL and sGFAP, respectively, before and after a relapse; the small vertical marks along the x-axis represent the individual samples. (B and D) Mean (dot) and 95% CI (error bars) of biomarkers over different time bins before, during, and after a relapse; *p* values are from LMM. Statistical significance is shown as *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. LMM = linear mixed-effects model; sGFAP = serum glial fibrillary acidic protein; sNfL = serum neurofilament light chain; sNfL = serum neurofilament light chain.

a subsequent clinical evaluation. We studied a series of sampling definition in which the length of the window determined whether a serum sample was classified as being collected during relapse or during remission (Figures 2 and 3). In a second step, we evaluated whether biomarker levels obtained within the selected relapse/remission definition windows can predict the risk of future EDA-3, analyzing protein Z-scores as both continuous and categorical thresholds at 1.0 and 1.5 (Figures 4 and 5). All analyses were performed using the full data set and then repeated in a sensitivity analysis excluding samples from patients treated with HE-DMTs.

Figure 2 Model Fit Metrics and Pairwise Comparison Across LMMs for Potential Prognostic Performance of sNfL

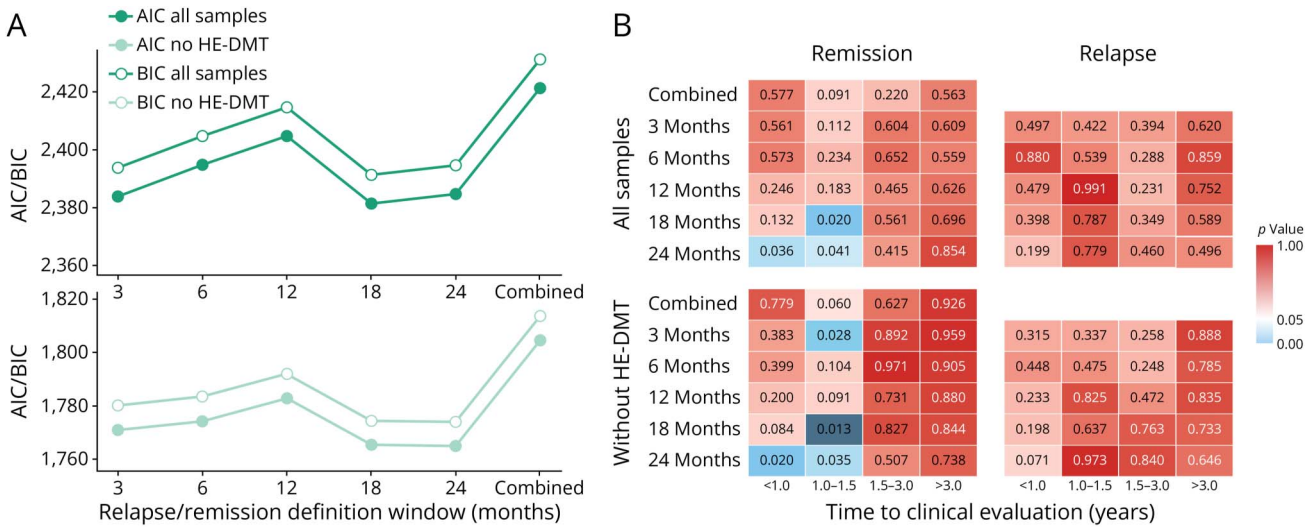


(A) Comparison of model fit indices (AIC and BIC) for LMMs. Dark purple represents results from the full cohort, while light purple corresponds to the cohort excluding samples on HE-DMTs. Filled circles indicate AIC values and empty circles indicate BIC values. (B) Heatmap summarizing *p* values from all LMMs comparing future NEDA-3 vs EDA-3 status. Colors range from dark blue (closest to 0) to dark red (1), with white representing the significance threshold ($p = 0.05$). AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; EDA-3 = evidence of disease activity-3; HE-DMT = high-efficacy disease-modifying therapy; LMM = linear mixed-effects model; NEDA-3 = no evidence of disease activity-3.

sNfL Z-Scores

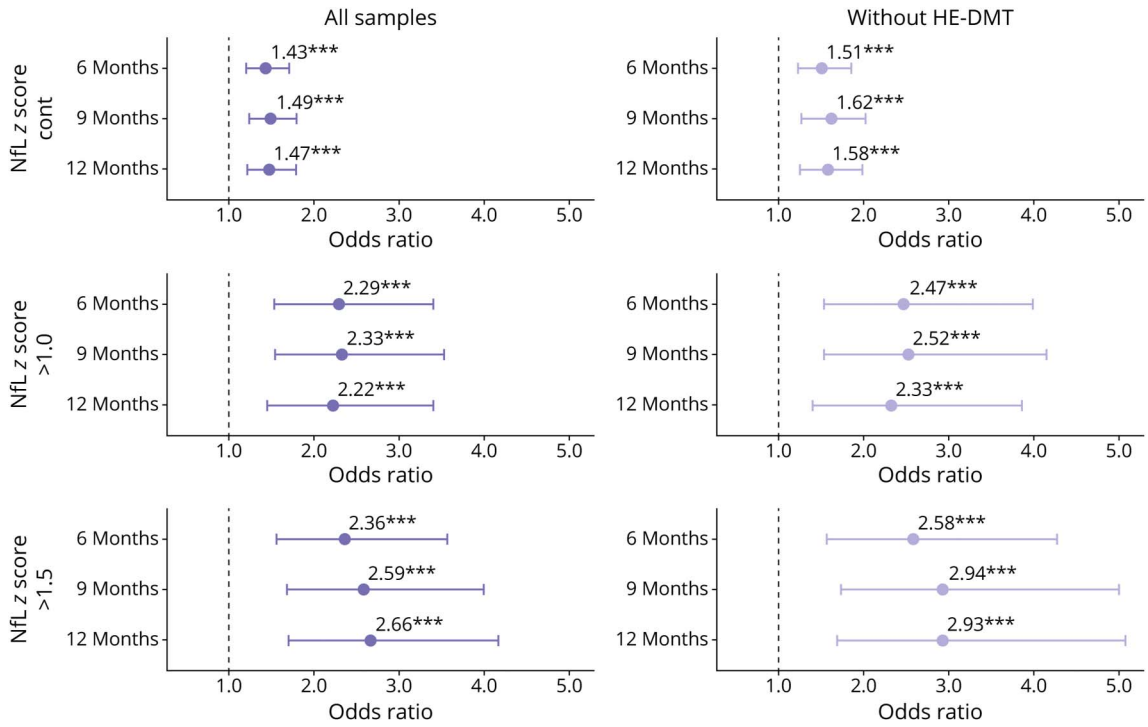
Using model fit metrics (AIC and BIC), we first evaluated how informative different relapse/remission definition windows were for distinguishing between future EDA-3 and NEDA-3 outcomes. In these models, the definition window determined whether serum samples were classified as relapse or remission related. As shown in Figure 2A, model fit improved when longer definition windows were used, indicating that very early postrelapse samples contribute limited prognostic information. However, models that combined relapse and remission samples consistently showed the poorest fit, supporting the biological distinction between these 2 clinical states and their separate evaluation. In addition, although extended windows (18–24 months) resulted in lower

Figure 3 Model Fit Metrics and Pairwise Comparison Across LMMs for Potential Prognostic Performance of sGFAP



(A) Comparison of model fit indices (AIC and BIC) for LMMs. Dark green represent results from the full cohort, while light green correspond to the cohort excluding patients on HE-DMTs. Filled circles indicate AIC values and open circles indicate BIC values. (B) Heatmap summarizing *p* values from all LMMs comparing future NEDA-3 vs EDA-3 status. Colors range from dark blue (closest to 0) to dark red (1), with white representing the significance threshold ($p = 0.05$). AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; EDA-3 = evidence of disease activity-3; LMM = linear mixed-effects model; NEDA-3 = no evidence of disease activity-3; sGFAP =serum glial fibrillary acidic protein.

Figure 4 Odds Ratios and CIs for EDA-3 Prognosis Using sNfL Z-Scores



OR with 95% CI for different prognostic models based on varying remission definitions (6, 9, and 12 months after a relapse). *p* values were derived from GLMMs with clinical outcome as the dependent variable and sNfL (continuous or categorized by thresholds) as the fixed effect. Dark purple bars represent analyses in the full cohort, while light purple bars correspond to analyses excluding patients on HE-DMTs. Statistical significance is indicated as ***$p < 0.001$. EDA-3 = evidence of disease activity-3; GLMM = generalized linear mixed model; HE-DMT = high-efficacy disease-modifying therapy; sNfL = serum neurofilament light chain.

AIC values, these windows increasingly classified stable remission samples as relapse related, reducing interpretability despite improved statistical fit. Therefore, window selection was guided not only by model fit but also by clinical and biological relevance.

Within remission samples, sNfL Z-scores were significantly higher in patients who later developed EDA-3 compared with those who remained NEDA-3, but only when samples were collected less than 1.5 years before the next clinical evaluation ($p < 0.001$; Figure 2B). These findings indicate that remission samples obtained approximately 6–12 months after relapse provide the earliest and most clinically interpretable separation of future disease activity and were therefore selected for subsequent prognostic analyses.

In the prognostic evaluation on the selected windows, higher sNfL Z-scores were strongly associated with increased risk of future EDA-3 within the following 1.5 years. OR ranged from 1.43 to 1.49 ($p < 0.001$). Categorizing sNfL Z-scores further enhanced prognostic discrimination. Using the 1.5 Z-score threshold produced the strongest effect, particularly in the 12-month window, where values above this cutoff were associated with an OR of 2.66 (95% CI 1.70–4.17). Higher thresholds consistently provided the clearest prognostication of future EDA-3 (Figure 4).

Sensitivity Analysis Excluding HE-DMT Samples

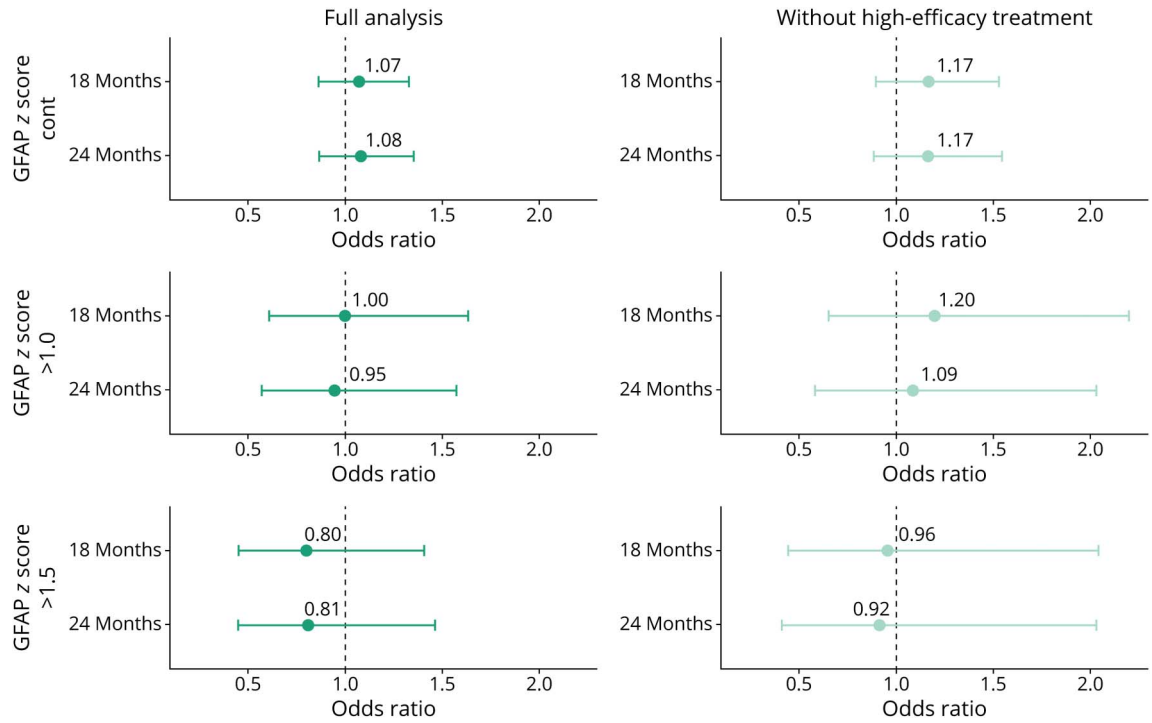
Repeating all analyses after excluding samples from patients undergoing HE-DMTs resulted in consistently improved model fit, likely reflecting reduced heterogeneity in biomarker responses. Despite this restriction, the overall pattern of results remained unchanged (Figure 2A).

Differences in sNfL Z-scores between patients who later developed EDA-3 and those who remained stable persisted across the same relapse/remission definitions and clinical evaluation intervals. The prognostic analysis became even stronger in the data set excluding HE-DMT; sNfL Z-scores as a continuous variable showed OR between 1.51 and 1.62. Categorized analyses again showed increased effect sizes, reaching an OR of 2.94 (95% CI 1.73–5.0) for the 1.5 cutoff in the 9-month relapse/remission definition (Figure 2B). The removal of HE-DMT samples strengthened the underlying associations.

sGFAP Z-Scores

The same analytical approach was applied to sGFAP Z-scores. In contrast to sNfL, the model fit showed only modest variation across relapse/remission definition windows, suggesting limited short-term temporal dynamics in this biomarker. Analyses that combined relapse and remission samples again performed worst, confirming that these states should be evaluated separately (Figure 3A).

Figure 5 Odds Ratios and CIs for EDA-3 Prognosis Using sGFAP



OR with 95% CI for different prognostic models based on varying remission definitions (18 and 24 months after a relapse). *p* values were derived from GLMMs with clinical outcome as the dependent variable and sNfL (continuous or categorized by thresholds) as the fixed effect. Dark green bars represent analyses in the full cohort, while light green bars correspond to analyses excluding patients on HE-DMTs. Statistical significance is indicated as $*p < 0.05$. EDA-3 = evidence of disease activity-3; GLMM = generalized linear mixed model; HE-DMT = high-efficacy disease-modifying therapy; sGFAP = serum glial fibrillary acidic protein; sNfL = serum neurofilament light chain.

Across most windows, sGFAP levels did not meaningfully differ between patients who later developed EDA-3 and those who remained NEDA-3. Only 2 extended remission windows (18- and 24-month definitions) showed significant differences when the next clinical evaluation was less than 1.5 years later (Figure 3B).

Within these relapse/remission definitions, sGFAP Z-scores showed no higher risk with future EDA-3, whether analyzed as a continuous value or using categorical thresholds (Figure 5). This indicates that sGFAP does not provide meaningful prognostic information for future EDA-3 in this short-term context.

Sensitivity Analysis Excluding HE-DMT Samples

Excluding samples obtained during HE-DMT modestly improved model fit, but the overall pattern remained unchanged (Figure 3A). Small increases in significance were observed in a few comparisons, but these did not translate into consistent differences across definitions (Figure 3B). Prognostic analyses again showed no significant associations between sGFAP levels and risk for future disease activity, regardless of model specification (Figure 5).

Discussion

The aim of this study was to characterize the temporal dynamics of sNfL and sGFAP in relation to clinical relapses and to evaluate their prognostic value for future EDA-3 in MS, with particular emphasis on their implications for routine clinical practice. We performed 2 complementary analyses: the first characterized biomarker trajectories around relapses and the second assessed prognostic performance across different sampling windows. Together, these findings intend to help refine guidelines for the appropriate clinical implementation of these blood-based biomarkers.

We observed a pronounced nonlinear trajectory of sNfL around relapses, with elevations detectable up to 1 year postrelapse and even 6 months before relapse onset. These findings are consistent with previous reports, indicating that sNfL increases during relapse periods and may rise before clinical onset and MRI activity.[11-13] A similar postrelapse dynamic was reported in alemtuzumab-treated patients, where sNfL peaked about 3 months after relapse before gradually normalizing.[16] The extended duration of elevated levels likely reflects ongoing acute axonal injury and delayed clearance of neurofilament proteins.[13,15] This wide dynamic window reinforces the utility of sNfL as a marker of inflammatory activity in the clinical context of MS and supports recommendations for frequent monitoring alongside the current imaging strategies.[17,30-32] This study further describes the optimal timing for prognostic sampling, demonstrating that remission samples collected 9 months postrelapse provide the clearest and earliest signal for predicting risk for

short-term disease activity. Because sNfL approaches its implementation in routine clinical practice, a precise understanding of its behavior across different phases of MS pathophysiology is critical to generate improved guidelines for individualized treatment decisions.[13-16]

In contrast to sNfL, sGFAP did not exhibit a strong or consistent temporal pattern; we found only a modest nonlinear association with relapse timing. Nevertheless, sGFAP elevations were observed at relapse and up to 2 years afterward, similar to findings reported in CSF.[8] These results suggest that astrocytic activation is not tightly related to acute inflammatory events but may occur in response to other pathologic processes, such as gliosis or chronic tissue remodeling. Previous work has linked sGFAP to progression independent of relapse activity,[18-22,33] supporting its role as a marker of neurodegenerative rather than short-term inflammatory processes. However, more research is needed to closely examine long-term sGFAP dynamics and to determine its value and optimal monitoring in individuals at risk for disease progression.

We also found remission sampling to be more informative than relapse sampling for predicting future disease activity. This likely reflects the confounding effect of acute inflammation during relapse, which elevates biomarker levels masking underlying trajectories. For sNfL, samples collected 9 months postrelapse yielded the strongest predictive performance for EDA-3 within the following 1.5 years, with odds ratios comparable with those reported by Benkert et al.[11] These findings highlight the importance of optimal sampling timing in blood-based biomarker monitoring. Although sGFAP demonstrated weaker prognostic utility in this context, its complementary role in capturing noninflammatory processes needs further investigation, particularly in progressive MS phenotypes.

Although sNfL moves closer to its integration into clinical practice for monitoring inflammatory activity, it represents only 1 facet of MS pathophysiology. To advance in precision medicine, incorporating complementary blood-based biomarkers that reflect additional biological mechanisms alongside current clinical and imaging-based monitoring remains essential.[30,32] Several studies have demonstrated that combining sNfL and sGFAP enhances clinical interpretation, improving early MS diagnosis,[34] informing treatment initiation,[35] and identifying patients at risk for either more active and/or progressive phenotypes.[16,22,26] To support standardized clinical use, confounding factors must be addressed; accordingly, we and others have used adjusted Z-scores, which offer a more robust basis for general standards and cutoff development.[27,36] These combined efforts are essential for building reliable decision algorithms and advancing individualized treatment.[22,26,35] Our analysis focused on the relationship between biomarker dynamics and clinical relapses. Although MRI is a valuable tool for disease monitoring of treatment, future studies should explore how blood-based biomarkers can be optimally integrated with the current imaging protocols to improve overall disease assessment.

The influence of HE-DMTs was clearly visible in our cohort. Prognostic performance improved when excluding samples from patients receiving these therapies; suggesting that because they profoundly modulate the immune system, the strong suppression of focal inflammation may also attenuate biomarker levels. Similar findings were observed in early disease, where HE-DMT exposure reduced the risk of future disease activity associated with high sNfL.[35] Treatment-specific associations with biomarker behavior have been documented in several studies,[22,35,37,38] and personalized treatment algorithms increasingly incorporate them.[30] These findings support the need to account for treatment context when interpreting biomarker levels in clinical practice.[30,31]

Strengths of this study include the use of a real-world patient cohort from a routine clinical setting, extensive longitudinal sampling, application of Z-scores to correct for confounders,[27] and robust statistical modeling to account for nonlinear dynamics and repeated measures. Limitations include a smaller sample size for modeling biomarker changes in the months preceding relapse, as well as the lack of distinction between relapses associated with Gd + lesions. In addition, although the cohort includes repeated longitudinal measurements, sampling intervals were not optimized to capture short-term biomarker dynamics around relapse onset; a more detailed characterization of acute sNfL changes would likely require more frequent sampling, such as monthly intervals. The heterogeneous composition of the cohort reflects real-world clinical practice but may limit the interpretability of subgroup-specific findings; therefore, future studies in larger, more homogeneous or stratified cohorts will be important to validate these findings and refine biomarker-based risk stratification. Moreover, these results should be further studied in combination with other fluid and imaging biomarkers.

In summary, this study demonstrates that sNfL follows a clear nonlinear temporal dynamic around relapses and provides strong short-term prognostic value, particularly when measured 9 months after a relapse. sGFAP shows discrete changes, suggesting a complementary role in capturing noninflammatory processes. Understanding the temporal dynamics of blood-based biomarkers and how the different treatments affect them is fundamental to their correct implementation in clinical settings. Optimal sampling strategies should prioritize remission periods to enhance predictive accuracy and guide personalized treatment decisions.

In conclusion, because these biomarkers approach clinical implementation, detailed characterization of their dynamics across clinical contexts becomes increasingly important. Our findings suggest that both sNfL and sGFAP should be monitored regularly and repeatedly, particularly during remission, to identify patients at risk of future disease activity.

Acknowledgment

The authors would like to express their gratitude to Mrs. Roberta Bichler for her invaluable contribution to the serum measurements of NfL and GFAP.

Author Contributions

M. Martinez-Serrat: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data. C. Tafrali: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data. R. Demjaha: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data. T. Kaiser: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design. M.T. Haindl: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data. B. Helmlinger: drafting/revision of the manuscript for content, including medical writing for content. E. Hofer: drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data. S. Ropele: major role in the acquisition of data. B. Heschl: major role in the acquisition of data. A. Damulina: major role in the acquisition of data. E. Platos: major role in the acquisition of data. D. Pinter: major role in the acquisition of data. F. Wohlrab: drafting/revision of the manuscript for content, including medical writing for content. T. Usnich: drafting/revision of the manuscript for content, including medical writing for content. F. Paul: drafting/revision of the manuscript for content, including medical writing for content. D. Leppert: major role in the acquisition of data. P. Benkert: major role in the acquisition of data. J. Kuhle: major role in the acquisition of data. C. Enzinger: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data. M. Khalil: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design.

Study Funding

The authors report no targeted funding.

Disclosure

The authors report no relevant disclosures. Go to Neurology.org/OA for full disclosures.

Publication History

Received by *Neurology® Open Access* March 31, 2026. Accepted in final form June 11, 2026. Submitted and externally peer reviewed. The handling editor was Editor Amy Kunchok, MBBS, MMed, FRACP, PhD.

References

1. Jakimovski D, Bittner S, Zivadinov R, et al. Multiple sclerosis. *Lancet.* 2024; 403(10422):183-202. doi:10.1016/S0140-6736(23)01473-3
2. Cao Y, Xu Y, Cao M, et al. Fluid-based biomarkers for neurodegenerative diseases. *Ageing Res Rev.* 2025;108:102739. doi:10.1016/J.ARR.2025.102739
3. Di Filippo M, Gaetani L, Centonze D, et al. Fluid biomarkers in multiple sclerosis: from current to future applications. *Lancet Reg Health Eur* 2024;44:101009. doi:10.1016/J.LANEPE.2024.101009
4. Chitnis T, Magliozzi R, Abdelhak A, Kuhle J, Leppert D, Bielekova B. Blood and CSF biomarkers for multiple sclerosis: emerging clinical applications. *Lancet Neurol.* 2025; 24(12):1066-1078. doi:10.1016/S1474-4422(25)00249-2
5. Khalil M, Teunissen CE, Lehmann S, et al. Neurofilaments as biomarkers in neurological disorders - towards clinical application. *Nat Rev Neurol.* 2024;20(5):269-287. doi:10.1038/S41582-024-00955-X
6. Bittner S, Oh J, Havrdová EK, Tintoré M, Zipp F. The potential of serum neurofilament as biomarker for multiple sclerosis. *Brain.* 2021;144(10):2954-2963. doi:10.1093/brain/awab241
7. Khalil M, Teunissen CE, Otto M, et al. Neurofilaments as biomarkers in neurological disorders. *Nat Rev Neurol.* 2018;14(10):577-589. doi:10.1038/s41582-018-0058-z
8. Malmeström C, Haghighi S, Rosengren L, Andersen O, Lycke J. Neurofilament light protein and glial fibrillary acidic protein as biological markers in MS. *Neurology.* 2003; 61(12):1720-1725. doi:10.1212/01.WNL.0000098880.19793.b6
9. Rosenstein I, Axelsson M, Novakova L, Blennow K, Zetterberg H, Lycke J. Exploring CSF neurofilament light as a biomarker for MS in clinical practice; a retrospective registry-based study. *Mult Scler J.* 2022;28(6):872-884. doi:10.1177/13524585211039104
10. Huss A, Senel M, Abdelhak A, et al. Longitudinal serum neurofilament levels of multiple sclerosis patients before and after treatment with first-line immunomodulatory therapies. *Biomedicines.* 2020;8(9):312. doi:10.3390/biomedicines8090312
11. Benkert P, Meier S, Schaedelin S, et al. Serum neurofilament light chain for individual prognostication of disease activity in people with multiple sclerosis: a retrospective modelling and validation study. *Lancet Neurol.* 2022;21(3):246-257. doi:10.1016/S1474-4422(22)00009-6
12. Thebault S, Reaume M, Marrie RA, et al. High or increasing serum NfL is predictive of impending multiple sclerosis relapses. *Mult Scler Relat Disord.* 2022;59:103535. doi:10.1016/J.MSARD.2022.103535
13. Akgün K, Kretschmann N, Haase R, et al. Profiling individual clinical responses by high-frequency serum neurofilament assessment in MS. *Neurol Neuroimmunol Neuroinflamm.* 2019;6(3):e555. doi:10.1212/NXI.0000000000000555
14. Fox RJ, Cree BAC, de Sèze J, et al. Temporal relationship between serum neurofilament light chain and radiologic disease activity in patients with multiple sclerosis. *Neurology.* 2024;102(9):e209357. doi:10.1212/WNL.0000000000209357
15. Rosso M, Gonzalez CT, Healy BC, et al. Temporal association of sNfL and gad-enhancing lesions in multiple sclerosis. *Ann Clin Transl Neurol.* 2020;7(6):945-955. doi:10.1002/acn3.51060
16. Inojosa H, Werder L, Haase R, et al. Longitudinal high-frequency blood biomarkers of axonal injury and astrocytic activation after immune reconstitution in multiple sclerosis. *J Neuroinflamm.* 2025;23(1):37. doi:10.1186/S12974-025-03674-2
17. Freedman MS, Gnanapavan S, Booth RA, et al. Guidance for use of neurofilament light chain as a cerebrospinal fluid and blood biomarker in multiple sclerosis management. *EBioMedicine.* 2024;101:104970. doi:10.1016/j.ebiom.2024.104970
18. Abdelhak A, Antweiler K, Kowarik MC, et al. Serum glial fibrillary acidic protein and disability progression in progressive multiple sclerosis. *Ann Clin Transl Neurol.* 2024; 11(2):477-485. doi:10.1002/acn3.51969
19. Barro C, Healy BC, Liu Y, et al. Serum GFAP and NfL levels differentiate subsequent progression and disease activity in patients with progressive multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm.* 2023;10(1):e200052. doi:10.1212/NXI.0000000000200052
20. Rosenstein I, Nordin A, Sabir H, et al. Association of serum glial fibrillary acidic protein with progression independent of relapse activity in multiple sclerosis. *J Neurol.* 2024;271(7):4412-4422. doi:10.1007/s00415-024-12389-y
21. Meier S, Willemse EAJ, Schaedelin S, et al. Serum glial fibrillary acidic protein compared with neurofilament light chain as a biomarker for disease progression in multiple sclerosis. *JAMA Neurol.* 2023;80(3):287-297. doi:10.1001/jamaneurol.2022.5250
22. Benkert P, Maleska Maceski A, Schaedelin S, et al. Serum glial fibrillary acidic protein and neurofilament light chain levels reflect different mechanisms of disease progression under B-Cell depleting treatment in multiple sclerosis. *Ann Neurol.* 2024; 97(1):104-115. doi:10.1002/ana.27096
23. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology.* 2014;83(3):278-286. doi:10.1212/WNL.0000000000000560
24. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology.* 1983;33(11):1444-1452. doi:10.1212/WNL.33.11.1444
25. Teunissen CE, Petzold A, Bennett JL, et al. A consensus protocol for the standardization of cerebrospinal fluid collection and biobanking. *Neurology.* 2009;73(22): 1914-1922. doi:10.1212/WNL.0B013E3181C47CC2
26. Maleska Maceski A, Benkert P, Einsiedler M, et al. GFAP and NfL as predictors of disease progression and relapse activity in fingolimod-treated multiple sclerosis. *Brain.* 2025;149(7):2295-2307. doi:10.1093/BRAIN/AWAF433
27. Einsiedler M, Maceski AM, Sandgren S, et al. Serum neurofilament light chain in multiple sclerosis: superiority of age- and BMI-Corrected Z scores/percentiles over absolute cutoff values for prediction of treatment response. *Ann Clin Transl Neurol.* 2025;12(11):2214-2225. doi:10.1002/ACN3.70149
28. Buchmann A, Pirpamer L, Pinter D, et al. High serum neurofilament light chain levels correlate with brain atrophy and physical disability in multiple sclerosis. *Eur J Neurol.* 2023;30(5):1389-1399. doi:10.1111/ENE.15742

29. Bsteh G, Marti S, Hammer H, et al. Dissecting definitions of disability accrual in relapsing multiple sclerosis—Have we reached standardization yet? *Mult Scler*. 2026; 32(2):179-191. doi:10.1177/13524585251396283
30. Yaldizli Ö, Benkert P, Achtnichts L, et al. Personalized treatment decision algorithms for the clinical application of serum neurofilament light chain in multiple sclerosis: a modified Delphi study. *Mult Scler*. 2025;31(8):932-943. doi:10.1177/13524585251335466
31. Bar-Or A, Nicholas J, Feng J, Sorrell F, Cascione M. Exploring the clinical utility of neurofilament light chain assays in multiple sclerosis management. *Neurol Neuroimmunology Neuroinflammation*. 2025;12(4):e200427. doi:10.1212/NXI.0000000000200427
32. Barkhof F, Reich DS, Oh J, et al. 2024 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI for the diagnosis of multiple sclerosis. *Lancet Neurol*. 2025;24(10):866-879. doi:10.1016/S1474-4422(25)00304-7
33. Abdelhak A, Bachhuber F, Ning K, et al. Blood biomarkers for predicting disability worsening in progressive multiple sclerosis: a multinational, individual participant-level analysis. *J Neurol Neurosurg Psychiatry* 2025;96(11):1046-1052. doi:10.1136/JNNP-2025-335831
34. Comabella M, Pappolla A, Monreal E, et al. Contribution of blood biomarkers to multiple sclerosis diagnosis. *Neurol Neuroimmunol Neuroinflamm*. 2025;12(2): e200370. doi:10.1212/NXI.0000000000200370
35. Monreal E, Fernández-Velasco JI, Álvarez-Lafuente R, et al. Serum biomarkers at disease onset for personalized therapy in multiple sclerosis. *Brain* 2024;147(12): 4084-4093. doi:10.1093/BRAIN/AWAE260
36. La Civita E, Nicolella V, Fiorenza M, et al. Advancing clinical use of neurofilament light chain: translational insights from research to routine practice. *Biomark Insights*. 2025;20:11772719251364018. doi:10.1177/11772719251364018
37. Bridel C, Leurs CE, Van Lierop ZYGJ, et al. Serum neurofilament light association with progression in natalizumab-treated patients with relapsing-remitting multiple sclerosis. *Neurology*. 2021;97(19):e1898-e1905. doi:10.1212/WNL.0000000000012752
38. Sainz-Amo R, Rodero Romero A, Monreal E, et al. Effect of alemtuzumab over sNfL and sGFAP levels in multiple sclerosis. *Front Immunol*. 2024;15:1454474. doi:10.3389/FIMMU.2024.1454474