



# Reliability of serum neurofilament light and glial fibrillary acidic protein for detecting disease activity upon discontinuation of first-line disease-modifying therapy in stable multiple sclerosis (DOT-MS)

W. H. Fung<sup>1</sup> · M. H. J. Wessels<sup>1</sup> · E. M. E. Coerver<sup>1</sup> · J. de Beukelaar<sup>2</sup> · W. H. Bouvy<sup>3</sup> · L. R. Canta<sup>4</sup> · O. H. H. Gerlach<sup>5</sup> · E. Hoitsma<sup>6</sup> · E. L. J. Hoogervorst<sup>7</sup> · B. A. de Jong<sup>1</sup> · N. F. Kalkers<sup>8</sup> · Z. L. E. van Kempen<sup>1</sup> · H. Lövenich<sup>9</sup> · C. E. P. van Munster<sup>10</sup> · B. W. van Oosten<sup>1</sup> · J. Smolders<sup>11</sup> · A. Vennegoor<sup>12</sup> · E. M. P. E. Zeinstra<sup>13</sup> · M. Barrantes-Cepas<sup>14</sup> · G. Kooij<sup>15</sup> · M. M. Schoonheim<sup>14</sup> · B. I. Lissenberg-Witte<sup>16</sup> · B. Moraal<sup>17</sup> · F. Barkhof<sup>17,18</sup> · B. M. J. Uitdehaag<sup>1</sup> · J. Mostert<sup>19</sup> · C. E. Teunissen<sup>20</sup> · J. Killestein<sup>1</sup> · E. M. M. Strijbis<sup>1</sup>

Received: 16 June 2025 / Revised: 23 June 2025 / Accepted: 24 June 2025  
© The Author(s) 2025

## Abstract

**Background** Neurofilament light(NfL) and glial fibrillary acidic protein(GFAP) are associated with disease activity in multiple sclerosis(MS), however use in monitoring remains limited. The ability of these biomarkers to detect disease activity upon treatment discontinuation was studied.

**Methods** Long-term stable relapse-onset MS patients were to continue or discontinue their first-line disease-modifying therapy(DMT) to study the safety of DMT discontinuation(DOT-MS trial NCT04260711). “Significant” disease activity was defined as clinical relapse,  $\geq 3$  new lesions or  $\geq 2$  contrast-enhancing lesions. MRI and sampling were performed at baseline, month 3, 6, 12, 18 and 24. Associations of delta biomarker levels and NfL z-score(age and body mass index derived) with “significant” disease activity were tested. Cut-off values for biomarkers to detect disease activity were calculated.

**Results** 45(50.5%) participants discontinued their DMT. Eight(all discontinued DMT) had “significant” disease activity, which was associated with an increase in NfL levels(OR:1.13 [1.03–1.33],  $p = 0.04$ ) and NfL z-scores(OR:2.17 [0.98–5.22],  $p = 0.06$ ), but not with GFAP( $p = 0.52$ ). Delta NfL had the highest ability to detect “significant” disease activity(AUC:0.88 [0.76–0.99]), with the best calculated cut-off of 46.4% increase(AUC:0.68, sensitivity 0.57, specificity 0.96).

**Discussion** NfL may be useful to identify, but not predict, disease activity after DMT discontinuation in MS. GFAP levels were not discriminatory for disease activity.

**Keywords** MS · MRI in MS · Clinical trials

## Introduction

The key pathological features behind the multiple sclerosis (MS) disease pathogenesis are thought to be neuroinflammation, demyelination and neurodegeneration [1, 2]. Blood-based biomarkers appear to be able to provide a relatively cheap and easily available method of assessing, monitoring or even predicting such processes on a group level [3–5]. However, implementation of these biomarkers in clinical practice remains slow [6].

Two widely investigated blood-based biomarkers in the context of MS are neurofilament light (NfL) and glial fibrillary acidic protein (GFAP). NfL and GFAP are proteins found in neurons and astrocytes, respectively. NfL serves as a specific biomarker for neuronal damage, with increasing levels during relapses or new MRI lesions [7], and decreasing levels upon effective treatment [8]. Additionally, NfL levels hold prognostic value for disability accumulation, especially in relapsing-remitting MS [9, 10]. On the other hand, GFAP is considered as a biomarker of astrocytic damage and reactive astrogliosis, although its association with acute disease activity is inconsistent in blood and cerebrospinal fluid samples [11, 12].

To use NfL and GFAP as tools for individual disease monitoring, standardized cut-off values still need to be

W. H. Fung and M. H. J. Wessels: shared co-first authorship.

Extended author information available on the last page of the article

established. Although fixed biomarker cut-offs for pathology and reference values correcting for age and Body Mass Index (BMI) have been explored previously, the individual use of NfL and GFAP has not been widely implemented as a clinically usable biomarker [13–16]. In our DOT-MS cohort (NCT04260711), we therefore investigated the ability of NfL and GFAP to discriminate disease activity and stable disease in participants with long-term stable MS who continued or discontinued first-line disease-modifying therapy (DMT).

## Methods

### Study design and participants

Eighty-nine participants were included from the DOT-MS trial [17]: an investigator-initiated, multicenter clinical trial, with randomized treatment-group assignments, open-label treatment and blinded end-point evaluation. Participants were randomized into continuation of DMT or discontinuation of DMT. Participants were 18 years or older, had relapse-onset MS, used first-line DMT (any of the interferons, glatiramer acetate, teriflunomide or dimethyl fumarate), and no clinical relapses nor substantial radiological disease activity (defined as no new contrast-enhancing lesions and  $\leq 1$  new T2 lesion on MRI) for at least 5 years before inclusion.

### Study procedures and outcomes

Participants underwent a clinical evaluation, brain MRI with gadolinium and blood sampling at baseline, 3, 6, 12, 18 and 24 months. At each visit, all assessments were performed on the same day. In addition to the scheduled visits, an unscheduled visit was arranged if deemed necessary by the physician. MRI assessments were performed by neuro-radiologists blinded to treatment allocation. In addition, scans that showed any MRI activity were centrally reviewed by a dedicated neuroradiologist in the main investigation center (Amsterdam UMC, location VUmc). Spinal cord imaging was only performed in relapses with suspected spinal cord involvement.

“Significant” disease activity was defined as any confirmed clinical relapse and/or “significant” MRI activity. “Significant” MRI activity was defined as  $\geq 3$  new T2 lesions or  $\geq 2$  contrast-enhancing lesions on brain MRI. In addition, we included “any” MRI activity, defined as “significant” disease activity and any other MRI activity (such as one or two new T2 lesions, one contrast-enhancing lesion or enlarged T2 lesion).

### Biomarker analysis

Blood samples were collected at baseline and all follow-up study visits. Blood was centrifuged at 1800 g for 10 minutes at room temperature, aliquoted, and stored at  $-80^{\circ}\text{C}$ . Serum NfL and GFAP levels were assessed using the Simoa™ N4PE kit on the Simoa™ HD-X instrument following the instructions (Quanterix, Billerica, USA) as described in detail previously [18–20]. Intra- and inter-assay precision of three quality control samples measured in duplicate over four runs ranged for NfL respectively from 7.0% to 8.7% and 7.0% to 9.8%, and for GFAP 4.2% to 12.2% and 7.7% to 12.2%.

We calculated absolute changes in NfL (absolute delta NfL) and GFAP (absolute delta GFAP) levels as the absolute difference between the NfL and GFAP levels in comparison to the previous measurement. Similarly, percentage changes in NfL and GFAP levels were calculated by determining the percentage difference between the NfL (percentage delta NfL) and GFAP (percentage delta GFAP) levels compared to those from the previous measurement. Lastly, for age and BMI-adjusted normative NfL values, z-scores were derived from the online application created by Benkert et al. [15] Absolute delta NfL levels were adjusted for BMI, except in the analysis involving cut-off values, where the absolute delta NfL levels were not BMI-adjusted.

### Statistical analysis

Statistical analysis was performed using R version 4.2.1. A  $p$  value  $\leq 0.05$  was considered statistically significant. Two-sample  $t$ -test and the Mann–Whitney test were used appropriately based on visual inspection of data distribution. Proportions were compared using the Chi-square test. To study the associations between delta NfL levels, delta GFAP levels and NfL z-score with “significant” disease activity and “any” MRI activity, we performed logistic regression analysis. The mean delta NfL levels, delta GFAP levels and NfL z-score at the time of first occurrence of “significant” disease activity or “any” MRI activity in those participants were compared with mean delta NfL levels, delta GFAP levels and NfL z-score over time in participants without such activity (Fig. 1A, B). Comparison of delta biomarker changes at first activity versus at last sampling in patients without activity are available as supplementary material. We calculated the odds of an event (“significant” disease activity or “any” MRI activity) at follow-up for delta NfL levels, delta GFAP levels and NfL z-score. Covariates (BMI, age, sex) were separately and only jointly entered into the delta NfL and delta GFAP

model, and the best fitting model was then selected based on the likelihood ratio tests. sNfL values were predicted with each model and compared to actual sNfL values using the Hosmer-Lemeshow test, after which a receiver operating characteristic (ROC) curve was used to assess sensitivity and specificity of each model for “significant” disease activity and “any” MRI activity. Cut-off values were selected from univariate models using Youden’s *J* statistic. The proposed cut-off value of the NfL *z*-score of  $>1.5$  for a significantly higher chance of clinical and/or radiological activity [15] was also considered. To evaluate the magnitude of NfL elevations, we calculated the 95th percentile threshold for percentage change from baseline in the participants who did not develop “significant” disease activity and “any” MRI activity. This threshold was then used to evaluate the magnitude of NfL elevations in participants who did develop disease activity.

In addition, we have investigated the association between GFAP changes and confirmed disability progression. Significant confirmed disability progression was defined as an increase of  $\geq 1.0$  point from the baseline EDSS score if the baseline score was  $\leq 5.5$  or an increase of  $\geq 0.5$  points if the baseline score was  $>5.5$ , sustained for at least 24 weeks. The change of GFAP levels for participants with confirmed disability progression was analyzed using 1-sample *t* test.

### Standard protocol approvals, registrations, and patient consents

Written informed consent was obtained from all participants. The use of data for this study was approved by the Medical Ethical Committee of the Amsterdam UMC, location VUmc (Protocol ID NL71260.029.19; clinicaltrials.gov identifier NCT04260711).

### Data availability

Anonymized data not published within this article will be made available by request from any qualified investigator.

## Results

Out of 89 participants ( $n = 44$  in the continue group and  $n = 45$  in the discontinue group), eight participants had “significant” disease activity, all in the discontinue group, as previously described [17]. There were four additional participants with “any” MRI activity other than “significant” disease activity. Thus, there were in total 12 participants with “any” MRI activity, one of which was in the continue group and 11 in the discontinue group. No participants experienced clinical relapse without inflammatory MRI activity.

The participants had a median follow-up of 15.3 months (IQR 11.4–23.9). Not all participants completed the 24-month follow-up due to premature termination of the trial: all cases of “significant” disease activity were observed in the discontinuation group, with no cases as such observed in the continuation group.

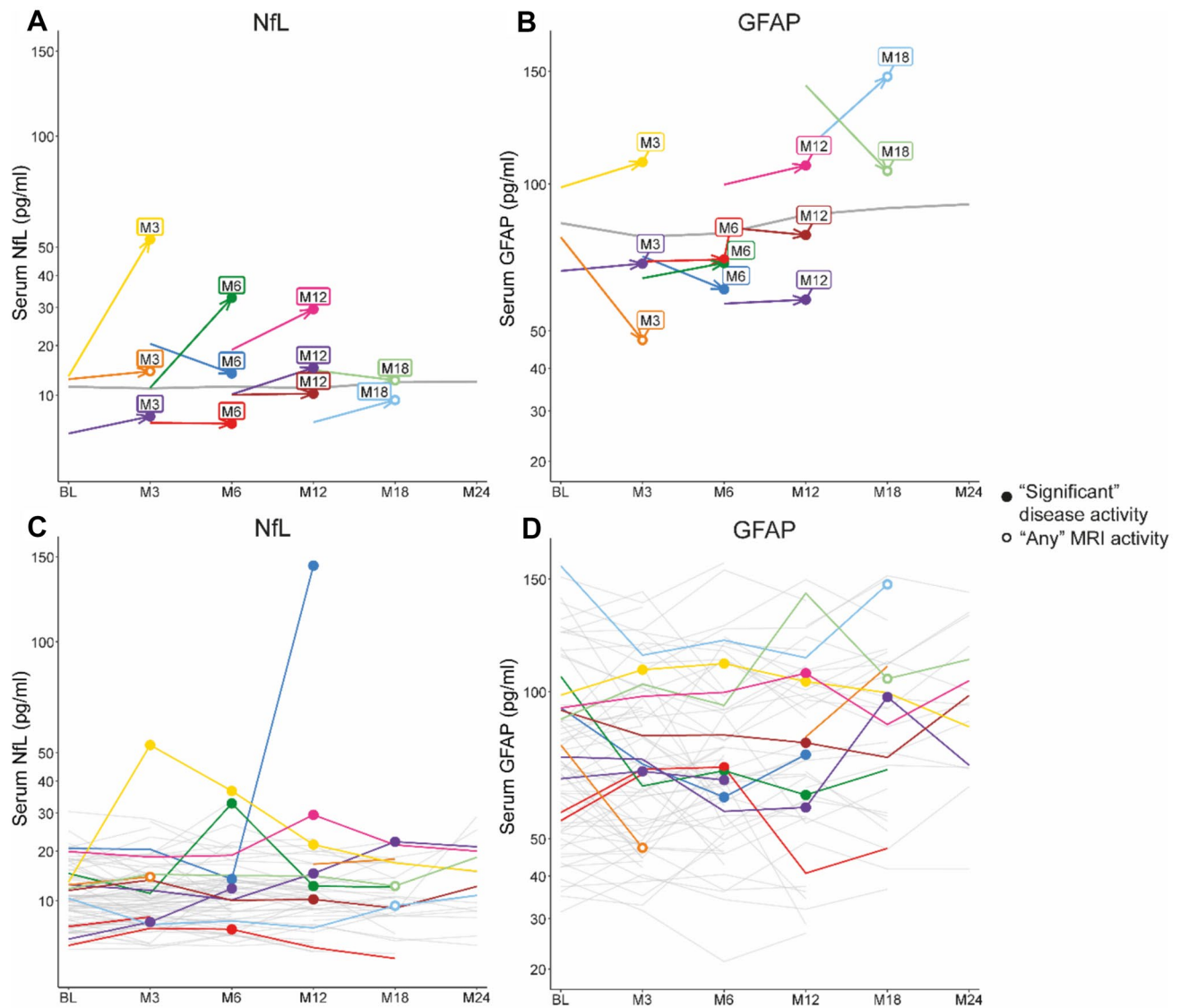
### Serum NfL and GFAP levels

Baseline NfL and GFAP levels did not differ between the continue and discontinue group (median NfL 10.53 [IQR 7.78–14.91] vs 10.27 [IQR 6.37–14.47],  $p = 0.59$ ; median GFAP 74.28 [IQR 53.67–109.62] vs 77.78 [IQR 58.55–94.94],  $p = 0.90$ ; Table 1). Also, no differences in baseline NfL or GFAP levels were found between participants with and without “significant” disease activity or “any” MRI activity (Table 1). Mean NfL levels of participants at visits during “significant” disease activity were higher than mean NfL levels at all visits of participants without “significant” disease activity (median NfL 18.31 [IQR 12.49–30.33] vs 10.37 [IQR 7.58–14.29],  $p < 0.001$ ). Additionally, mean NfL levels of participants at visits during “any” MRI activity were higher compared to mean NfL levels at all visits of participants without “any” MRI activity (median NfL level 13.94 [IQR 12.14–29.50] vs 10.27 [IQR 7.56–14.01],  $p = 0.002$ ). For the NfL change compared to the previous measurement, participants with “significant” disease activity or “any” MRI activity had a higher increase in their NfL level at the first occurrence of disease activity relative to the prior measurement compared to the participants without disease activity (Table 2). Absolute GFAP levels nor changes in GFAP were different between participants with or without “significant” disease activity or “any” MRI activity. Fig. 1C, D shows the longitudinal trajectories of NfL and GFAP levels per participant.

Absolute delta NfL levels, corrected for BMI (OR: 1.13, 95% CI [1.03–1.33],  $p = 0.04$ ), and percentage delta NfL levels (OR: 1.02, 95% CI [1.01–1.03],  $p = 0.01$ ) were associated with “significant” disease activity. For “any” MRI activity, there was a significant association observed with percentage delta NfL levels (OR: 1.02, 95% CI [1.01–1.03],  $p = 0.01$ ) and NfL *z*-score (OR: 2.08, 95% CI [1.07–4.30],  $p = 0.04$ ) but not with absolute delta NfL levels. In contrast, neither absolute delta GFAP levels, or percentage delta GFAP levels were significantly associated with “significant” disease activity or “any” MRI activity (Table 3).

### Sensitivity and specificity analyses

ROC curves were calculated for absolute delta NfL and GFAP levels, percentage delta NfL and GFAP levels, and NfL *z*-score for both “significant” disease activity” and



**Fig. 1** **A, B** The mean change in NfL and GFAP at time of first occurrence of "significant" disease activity or "any" MRI activity. The grey line represents the overall mean change in NfL and GFAP

levels in participants without such disease activity. **C, D** Individual longitudinal trajectories of NfL and GFAP levels in MS. Absolute serum NfL and GFAP levels are presented

"any" MRI activity. The model with absolute delta NfL change corrected for BMI performed best compared to percentage delta NfL or NfL z-score at the first moment of inflammatory disease activity, both in detecting "significant" disease activity (AUC 0.88, 95% CI [0.76–0.99]) and "any" MRI activity (AUC 0.83, 95% CI [0.70–0.97]) (Figure 2). In contrast, absolute delta GFAP levels and percentage delta GFAP level change were not able to detect "significant" disease activity or "any" MRI activity.

### Serum NfL cut-off value for inflammatory disease event

Univariate and uncorrected cut-off values for NfL levels in this cohort were calculated using Youden's J statistic (Table 2). For "significant" disease activity, a cut-off was calculated in percentage increase in NfL level ( $\geq 46.4\%$ , AUC 0.68, 95% CI [0.38–0.98], sensitivity 0.57, specificity 0.96, positive predictive value 0.57, negative predictive value 0.96). For "any" MRI activity, a cut-off was calculated in percentage increase in NfL level ( $\geq 46.4\%$ , AUC 0.70, 95% CI [0.47–0.93], sensitivity



0.55, specificity 0.96, positive predictive value 0.67, negative predictive value 0.93) and NfL *z*-score at moment of activity ( $\geq 2.33$ , AUC 0.63, 95% CI [0.38–0.87], sensitivity 0.27, specificity 1.00, positive predictive value 1, negative predictive value 0.89). A previously proposed cut-off of *Z*-score  $\geq 1.5$  [15] performed comparably (AUC 0.63, sensitivity 0.27, specificity 0.90, positive predictive value 0.3, negative predictive value 0.89).

### Changes in NfL levels prior first occurrence of disease activity

The 95th percentile threshold of percentage change from baseline was 46.47% for “significant” MRI activity. For “any” MRI activity, the 95th percentile threshold of percentage change from baseline was 45.85%. Only 1/8 participants with “significant” disease activity and 1/12 participants with “any” MRI activity exceeded the 95th percentile threshold for percentage change from baseline.

### GFAP and confirmed disability progression

There were no associations between GFAP levels and clinical disability. The change of GFAP levels between baseline and confirmed disability progression was 0.72% (SD 9.28;  $p = 0.83$ ), and the change between baseline and the visit prior confirmed disability progression was  $-2.45\%$  (SD 11.98;  $p = 0.58$ ).

## Discussion

Serum neurofilament light is known to be associated with disease activity, and has been used and proposed as a treatment response biomarker [7, 8]. Investigation into the clinical application of biomarkers in context of de-escalation is ongoing [21, 22]. In this study, we aimed to evaluate the utility of serum NfL and GFAP levels to discriminate participants with disease activity and stable disease. In the DOT-MS cohort, participants with “significant” disease activity and “any” MRI activity showed a clear increase in their NfL levels, while this was not apparent for GFAP levels. We did not see a difference in biomarker levels before disease activity. Based on the models’ performance, NfL levels were useful to detect disease activity due to high specificity, while GFAP levels were not discriminative of disease activity or stable disease.

Absolute change of serum NfL, corrected for BMI, best detected “significant” disease activity and “any” MRI activity at that timepoint, with the highest performance reached for “significant” disease activity. This could possibly be explained by higher NfL levels in more active disease states as it has been shown that higher serum NfL

levels correlate with the number and volume of contrast-enhancing lesions, and T2 lesion volume [23]. It should be noted that in all models involving NfL levels, the sensitivity for detecting a disease event is relatively low compared to a higher specificity. NfL could detect disease activity after DMT discontinuation and identify disease activity when a clinical relapse is suspected. This could help early recognition of the recurrence of significant inflammatory disease activity. However, NfL is less useful to rule out disease activity, and, therefore, clinical and radiological monitoring after DMT discontinuation is still necessary. In line with a previous study, NfL also had somewhat lower sensitivity and higher specificity to detect contrast-enhancing lesions in people with MS who continued or discontinued natalizumab. Next to that, the model for disease activity, including sNfL *z*-scores previously calculated underperformed in this cohort [15]. This could imply that standardized NfL *z*-scores may not universally apply to all individuals, while individual changes in NfL may be more indicative of disease activity. Moreover, the consideration of assay variability should be factored into the utilization of NfL *z*-score as a tool to detect disease activity. In addition, we have now defined criteria for “significant” disease activity and “any” MRI activity, but future research should also consider further categorization of disease activity into new T2 lesions, contrast-enhancing lesions, enlarging T2 lesions and clinical relapses, while also investigating the role of NfL in this context.

In the DOT-MS trial, we highlight that serum biomarkers (NfL and GFAP) did not predict subsequent later disease activity, consistent with findings from a prior study examining the temporal association between NfL levels and new disease activity [24]. Similarly, that study reported low sensitivity but high specificity of NfL in detecting disease activity, aligning with our results. In contrast, another study has shown that either NfL or GFAP following treatment discontinuation were associated with an increased risk of 6-month confirmed disability worsening and the development of new MRI activity [22]. While we also found a significant association between increased NfL levels and new disease activity, this relationship was not observed for GFAP. A possible explanation for this discrepancy is the difference in study design: the referenced study collected biomarker samples within two years following treatment discontinuation and assessed outcomes at follow-up several years later. In contrast, our analysis focused on the association between biomarker levels and new disease activity over a shorter period, within a 3–6-month time window. Additionally, the referenced study identified associations between biomarker levels and both disease activity (median 5.15 years after discontinuation) and confirmed disability progression (median 1.99 years after discontinuation). Unfortunately, our follow-up period is not yet long enough to capture such long-term

**Table 1.** Baseline characteristics of the study cohort

	All participants ( <i>n</i> = 89)	“Significant” disease activity ( <i>n</i> = 8)	No “significant” disease activity ( <i>n</i> = 81)	“Any” MRI activity ( <i>n</i> = 12)	No “any” MRI activity ( <i>n</i> = 77)
Median age (IQR), years	54 (49.0–59.0)	46 (43.5–58.5)	54 (50.0–59.0)	50 (43.5–60.25)	54 (50.0–58.0)
Sex					
Female	60 (67.4%)	5 (62.5%)	55 (68.0%)	9 (75.0%)	51 (66.2%)
Male	29 (32.6%)	3 (37.5%)	26 (32.0%)	3 (25.0%)	26 (33.8%)
Median time since symptom onset (IQR), years	14.0 (9.9–21.5)	14.5 (13.2–16.4)	14.0 (9.5–21.6)	13.5 (11.2–15.5)	14.2 (9.9–23.3)
Median time since last documented relapse (IQR), years	9.4 (6.9–13.2)	9.6 (7.4–11.2)	9.5 (6.9–13.2)	10.4 (8.1–12.3)	9.2 (8.1–12.3)
Multiple sclerosis subtype					
Relapsing–remitting	80 (89.9%)	8 (100.0%)	72 (88.9%)	12 (100.0%)	68 (88.3%)
Secondary progressive	9 (10.1%)	0 (0.0%)	9 (11.1)	0 (0.0%)	9 (11.7%)
Median total duration of disease-modifying therapy use (IQR), years	11.2 (7.7–16.0)	10.3 (8.8–11.4)	11.4 (7.7–16.2)	10.4 (8.1–12.3)	9.4 (6.8–13.2)
Disease-modifying therapy at randomisation					
Interferon beta	35 (39.3%)	2 (25.0%)	33 (40.7%)	3 (25.0%)	32 (41.5%)
Glatiramer acetate	23 (25.8%)	1 (12.5%)	22 (27.2%)	3 (25.0%)	20 (26.0%)
Teriflunomide	12 (13.5%)	2 (25.0%)	10 (12.3%)	2 (16.7%)	10 (13.0%)
Dimethyl fumarate	19 (21.3%)	3 (27.5%)	16 (19.8%)	4 (33.3%)	15 (19.5%)
Expanded Disability Status Scale score	3.1 (1.8)	3.3 (0.9)	3.1 (1.8)	2.9 (1.0)	3.1 (1.9)
Median neurofilament light chain (IQR), pg/mL	10.3 (7.1–14.9)	13.1 (10.0–16.3)	10.1 (7.1–14.4)	12.5 (9.3–13.8)	10.0 (7.1–14.9)
Median glial fibrillary acidic protein (IQR), pg/mL	75.0 (55.8–97.7)	93.2 (73.8–94.9)	73.6 (54.4–109.6)	91.1 (73.8–94.9)	72.4 (53.7–109.6)
Median follow-up (IQR), months	15.3 (11.4–23.9)	21.5 (18.0–24.1)	13.8 (10.3–23.7)	21.5 (18.0–24.1)	13.6 (9.2–23.7)
Median follow-up (IQR) till disease activity, months		12.0 (6.75–12.0)		12.0 (6.0–12.0)	

Data are *n* (%) or mean (SD), unless otherwise specified. “Significant” disease activity was defined as any confirmed relapse and/or “significant” MRI activity ( $\geq 3$  new T2 lesions or  $\geq 2$  contrast-enhancing lesions on brain MRI). “Any” MRI activity was defined as any new T2 lesions, contrast-enhancing lesions or enlarged T2 lesions in addition “significant” disease activity

**Table 2.** Change in absolute serum NfL and GFAP levels at follow-up

	All participants ( <i>n</i> = 89)	No “significant” disease activity ( <i>n</i> = 81)	No “any” MRI activity ( <i>n</i> = 77)
Median delta NfL (IQR), pg/mL	0.05 (−0.61–0.62)	0.02 (−0.63–0.52)	0.01 (−0.67–0.52)
Median delta GFAP (IQR), pg/mL	−0.43 (−3.61–3.80)	−0.43 (−3.61–3.80)	−0.43 (−3.87–3.58)
Median NfL z-score (IQR)	0.32 (−0.35–0.99)	0.32 (−0.36–0.96)	0.32 (−0.39–0.94)
		“Significant” disease activity ( <i>n</i> = 8)	“Any” MRI activity ( <i>n</i> = 12)
Median delta NfL during disease activity (IQR), pg/mL		2.14 (0.03–13.3)	2.14 (0.03–7.72)
Median delta GFAP during disease activity (IQR), pg/mL		1.20 (−1.07–3.74)	1.20 (−6.82–6.41)
Median NfL z-score during disease activity		0.64 (0.09–1.87)	0.64 (0.14–1.64)

Delta NfL and delta GFAP were calculated as the absolute change in NfL and GFAP levels compared to the previous measurement during the whole follow-up. For “significant” disease activity and “any” MRI activity, the delta NfL and delta GFAP were calculated as the absolute change in NfL and GFAP levels compared to the previous measurement at the time of first occurrence of disease activity

*NfL* neurofilament light chain, *GFAP* glial fibrillary acidic protein

**Table 3.** Logistic regression and sensitivity and specificity of NfL and GFAP

	Odds ratio [95% CI]	<i>p</i> -value*	AUC [95% CI]	Cut-off**	Sensitivity	Specificity
NfL in “significant” disease activity						
Absolute delta	1.13 [1.03–1.33]	0.04	0.88 [0.76–0.99]	≥2.14	0.57	0.89
Percentage delta	1.02 [1.01–1.03]	0.01	0.68 [0.38–0.98]	≥46.4	0.57	0.96
Z-score	2.17 [0.98–5.22]	0.06	0.63 [0.38–0.87]	≥2.79	0.29	1.00
NfL in “any” MRI activity						
Absolute delta	1.13 [1.02–1.33]	0.06	0.83 [0.70–0.97]	≥1.55	0.64	0.87
Percentage delta	1.02 [1.01–1.03]	0.01	0.70 [0.47–0.93]	≥46.4	0.55	0.96
Z-score	2.08 [1.07–4.30]	0.04	0.63 [0.44–0.81]	≥2.33	0.27	1.00
GFAP in “significant” disease activity						
Absolute delta	1.04 [0.95–1.20]	0.52	0.57 [0.34–0.80]			
Percentage delta	1.00 [0.94–1.08]	0.99	0.51 [0.29–0.74]			
GFAP in “any” MRI activity						
Absolute delta	1.00 [0.95–1.06]	0.90	0.55 [0.32–0.78]			
Percentage delta	0.98 [0.94–1.04]	0.53	0.50 [0.28–0.73]			

“Significant” disease activity was defined as any confirmed relapse and/or “significant” MRI activity (≥3 new T2 lesions or ≥2 contrast-enhancing lesions on brain MRI). “Any” MRI activity was defined as any new T2 lesions, contrast-enhancing lesions or enlarged T2 lesions in addition “significant” disease activity

\* *p*-value from logistic regression

\*\*Cut-off values were calculated univariate using Youden’s J statistic and not corrected for BMI

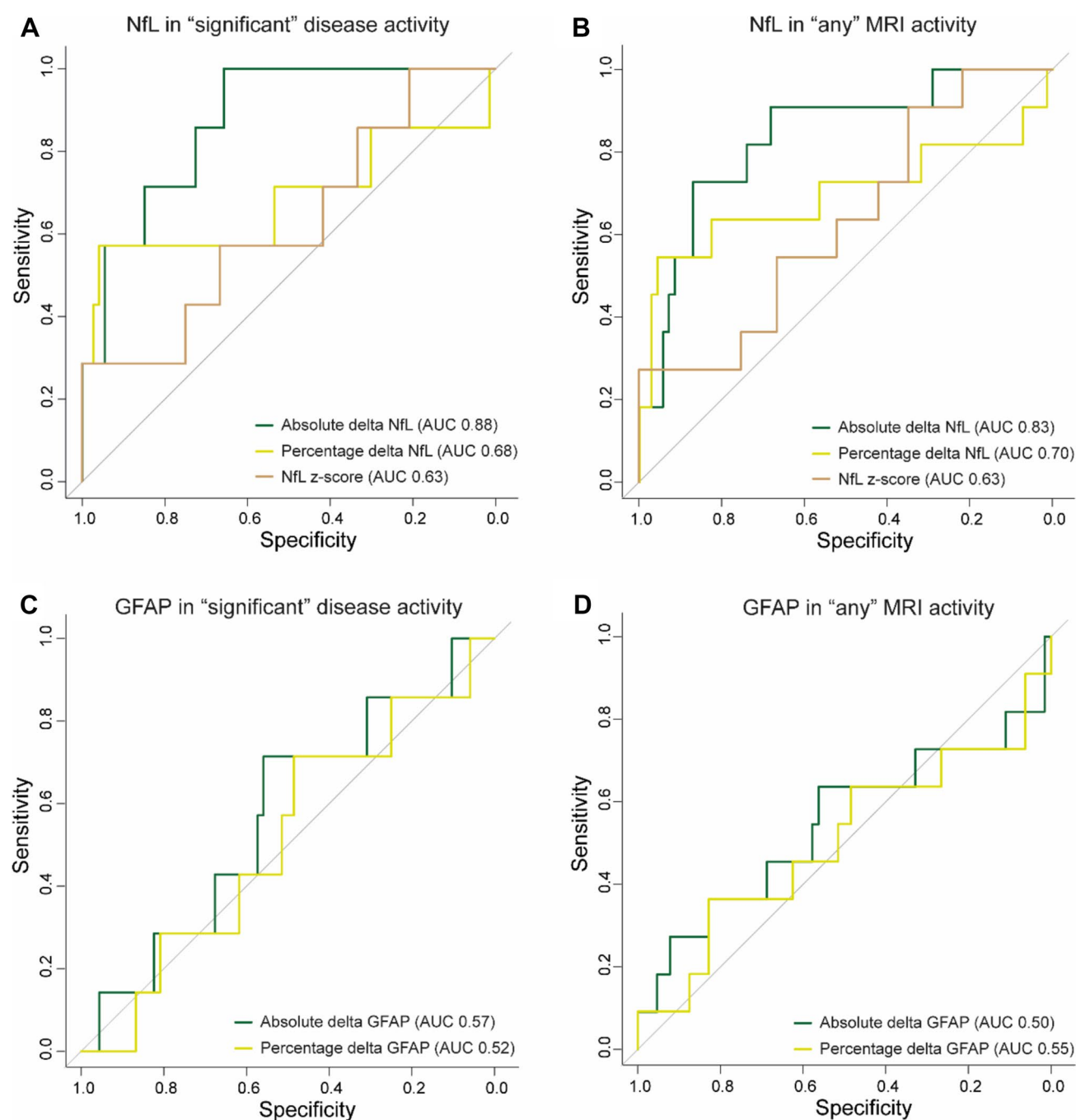
NfL neurofilament light chain, GFAP glial fibrillary acidic protein, AUC area under the Curve

outcomes. However, as the DOT-MS trial continues in an observational phase for an additional two years, future analyses will also focus on the association between biomarker levels and disease recurrence and confirmed disability progression at a longer follow-up duration.

Our study has limitations. Due to the early termination of the trial, the sample size of the DOT-MS trial was smaller than calculated in the preliminary power calculation: 45 and 45 versus the necessary sample size of 54 per group to achieve 80% power [17]. This resulted in a limited number of events. However, despite this limitation, we were still able to demonstrate the ability of NfL in identifying disease activity. Additionally, since blood sampling and MRI scans were conducted on the same day of the visit, assessing the predictive value of NfL for disease activity is challenging. To address this, we calculated the 95th percentile threshold for the percentage change from baseline in participants who did not develop disease activity, allowing us to evaluate the magnitude of NfL level elevations [23]. Factors such as renal function and other comorbidities, known to influence NfL levels, were not accounted for in the sensitivity and specificity analyses [6]. In addition, NfL is also a non-specific MS marker, and may indicate other neuronal pathology. Our analyses were

not corrected for the randomization group (continuation vs discontinuation of DMT) of the participants as the longitudinal NfL levels did not differ between discontinuation and continuation of DMT [17]. We suggest future research exploring NfL and its cut-offs further in a bigger cohort to validate our current findings. The DOT-MS cohort had a relatively older median age and remained stable over a long period. Future studies focusing on the NfL and GFAP course in the younger population are still needed.

In summary, absolute change of NfL levels corrected for BMI could serve as valuable tool for identifying and monitoring, but not predicting, disease activity in treatment discontinuation in MS. For now it remains a tool with caveats: the time between activity and NfL peaks is uncertain, NfL misses clear cut-off values for one-off measurements and clinical experience is still in its infancy [21, 24, 25]. While NfL levels alone may not be sufficient, the combination of NfL with clinical and MRI could enhance the ability and efficiency of identifying disease activity. Repeated NfL measurements could thus be used in adjunct to the necessary close clinical and radiological monitoring after discontinuation of a patient’s DMT. This data contributes to the evaluation of NfL’s role in regular MS care.



**Fig. 2** Receiver operating characteristic curves for NfL and GFAP in those with “significant” disease activity and “any” MRI activity. Absolute delta NfL models were corrected for BMI

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00415-025-13231-9>.

**Acknowledgements** This trial was kindly funded by ZonMW, grant number 848043001, and Stichting MS Research, grant number 17-992.

**Author contributions** Wessels and Fung shared first authorship. Wessels and Fung had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Wessels, Fung, Lissenberg-Witte, Killestein, Strijbis. Acquisition, analysis, or interpretation of data: Wessels, Fung, Coerver, de Beukelaar, Bouvy, Canta, Gerlach, Hoitsma, Hoogervorst, de Jong, Kalkers, van Kempen, Lövenich, van Munster, van Oosten, Smolders, Vennegoor, Zeinstra, Barrantes-Cepas, Schoonheim, Lissenberg-Witte, Teunissen, Moraal, Barkhof, Uitdehaag, Mostert, Killestein, Strijbis. Drafting of the manuscript: Wessels, Fung, Teunissen, Killestein, Strijbis. Critical review of the manuscript for important intellectual content: All authors. Statistical analysis: Wessels, Fung,



Lissenberg-Witte, Strijbis. Obtained funding: Killestein, Strijbis. Administrative, technical, or material support: Wessels, Fung, de Beukelaar, Gerlach, Hoitsma, Kalkers, Lövenich, Smolders, Zeinstra, Barrantes-Cepas, Kooji, Schoonheim, Teunissen, Moraal, Barkhof, Strijbis. Supervision: Teunissen, Killestein, Strijbis.

**Funding** This work was supported by ZonMW, 848043001, Stichting MS Research, 17-992.

## Declarations

**Conflicts of interest** W.H. Fung: nothing to report; M. Wessels: nothing to report; E.M.E. Coerver: nothing to report; J. de Beukelaar received (research) grants from the Dutch National MS Foundation (OZ2019-19, OZ2021-19 and P2023-003) and Stichting BeterKeten; W.H. Bouvy: nothing to report; L. Canta received study compensation fees for Consonance study, Esteem study, Lemtrada Pass study and TOP study; O.H.H. Gerlach: nothing to report; E. Hoitsma received speaker fees from Biogen, Roche, Sanofi Genzyme; authorship fee from Sandoz and study compensation fees for Calliper study, Esteem study and Lemtrada Pass study (study payments to institution); E. Hoogervorst: nothing to report; B.A. de Jong receives research support from Dutch MS Research Foundation, National MS Foundation, Dutch Multiple Sclerosis Association, and Zorgverzekeraars Nederland (i.e. umbrella organization of health insurers in the Netherlands); N.F. Kalkers: nothing to report; Z.L.E. van Kempen: nothing to report; H. Lövenich: nothing to report; C.E.P. van Munster: nothing to report; B.W. van Oosten: nothing to report; J. Smolders received speaker and/or consultancy fees from Biogen, Merck, Novartis, Roche and Sanofi; A. Vennegoor: nothing to report; E. Zeinstra served on advisory boards/received consultancy fees for Merck, Novartis, Genzyme and Roche; M. Barrantes-Cepas is supported by research grants from Merck and Atara Biotherapeutics; G. Kooij received grant from the Dutch Research Council (NWO Vidi grant 91719305 to G.K.), the Dutch MS Research Foundation (18-1023MS to G.K.) and research grants from Biogen, Merck and Novartis; M.M. Schoonheim serves on the editorial board of *Neurology*, *Multiple Sclerosis Journal* and *Frontiers in Neurology*, receives research support from the Dutch MS Research Foundation, Eurostars-EUREKA, ARSEP, Amsterdam Neuroscience, MAGNIMS and ZonMW (Vidi grant, project number 09150172010056) and has served as a consultant for or received research support from Atara Biotherapeutics, Biogen, Celgene/Bristol Meyers Squibb, EIP, Sanofi, MedDay and Merck; B.I. Lissenberg-Witte: nothing to report; B. Moraal: nothing to report; F. Barkhof serves on the steering committee and is iDMC member for Biogen, Merck, Roche, Eisai. He acts as a consultant for Roche, Biogen, Merck, IXICO, Janssen, Combinostics. He has research agreements with Novartis, Merck, Biogen, GE, Roche. He is co-founder and share-holder of Queen Square Analytics LTD; B.M.J. Uitdehaag reports consultancy fees from Immunic Therapeutics; J. Mostert: nothing to report; C.E. Teunissen reports funding from National MS Society (Progressive MS alliance) and Innovative Medicines Initiatives 3TR (Horizon 2020, grant no 831434). CET has a research contract with Celgene. She serves on editorial boards of *Medidact Neurologie/Springer*, *Neurology: Neuroimmunology & Neuroinflammation*, and is editor of a *Neuromethods* book Springer; J. Killestein received research grants for multicentre investigator initiated trials DOT-MS trial, ClinicalTrials.gov Identifier: NCT04260711 (ZonMW) and BLOOMS trial (ZonMW and Treatmeds), ClinicalTrials.gov Identifier: NCT05296161; received consulting fees for F. Hoffmann-La Roche, Biogen, Teva, Merck, Novartis and Sanofi/Genzyme (all payments to institution); reports speaker relationships with F. Hoffmann-La Roche, Biogen, Immunic, Teva, Merck, Novartis and Sanofi/Genzyme (all payments to institution); adjudication committee of MS clinical trial of Immunic (payments to institution only); E.M.M. Strijbis serves on the editorial board of *Frontiers in Neurology* and

receives research support from Stichting MS Research and ZonMW. Received speaker fees from Merck and Novartis.


**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Giovannoni F, Quintana FJ (2020) The role of astrocytes in CNS inflammation. *Trends Immunol* 41:805–819
2. Mahad DH, Trapp BD, Lassmann H (2015) Progressive multiple sclerosis I pathological mechanisms in progressive multiple sclerosis. *Lancet Neurol* 14:183–193
3. Teunissen CE, Malekzadeh A, Leurs C, Bridel C, Killestein J (2015) Body fluid biomarkers for multiple sclerosis—the long road to clinical application. *Nat Rev Neurol* 11:585–596
4. Novakova L, Zetterberg H, Sundstrom P et al (2017) Monitoring disease activity in multiple sclerosis using serum neurofilament light protein. *Neurology* 89:2230–2237
5. Abdelhak A, Foschi M, Abu-Rumeileh S et al (2022) Blood GFAP as an emerging biomarker in brain and spinal cord disorders. *Nat Rev Neurol* 18:158–172
6. Thebault S, Booth RA, Rush CA, MacLean H, Freedman MS (2021) Serum neurofilament light chain measurement in MS: hurdles to clinical translation. *Front Neurosci* 15:654942
7. Disanto G, Barro C, Benkert P et al (2017) Serum neurofilament light: a biomarker of neuronal damage in multiple sclerosis. *Ann Neurol* 81:857–870
8. Kuhle J, Kropshofer H, Haering DA et al (2019) Blood neurofilament light chain as a biomarker of MS disease activity and treatment response. *Neurology* 92:E1007–E1015
9. Thebault S, Abdoli M, Fereshtehnejad SM, Tessier D, Tabard-Cossa V, Freedman MS (2020) Serum neurofilament light chain predicts long term clinical outcomes in multiple sclerosis. *Sci Rep* 10:10381
10. Freedman MS, Abdelhak A, Bhutani MK et al (2025) The role of serum neurofilament light (sNfL) as a biomarker in multiple sclerosis: insights from a systematic review. *J Neurol* 272:400
11. Abdelhak A, Huss A, Kassubek J, Tuman H, Otto M (2018) Serum GFAP as a biomarker for disease severity in multiple sclerosis. *Sci Rep* 8:14798
12. Kassubek R, Gorges M, Schocke M et al (2017) GFAP in early multiple sclerosis: a biomarker for inflammation. *Neurosci Lett* 657:166–170
13. Hermesdorf M, Leppert D, Maceski A et al (2022) Longitudinal analyses of serum neurofilament light and associations with obesity indices and bioelectrical impedance parameters. *Sci Rep* 12:15863
14. Koini M, Pirpamer L, Hofer E et al (2021) Factors influencing serum neurofilament light chain levels in normal aging. *Aging (Albany NY)* 13:25729–25738
15. Benkert P, Meier S, Schaedelin S et al (2022) Serum neurofilament light chain for individual prognostication of disease

- activity in people with multiple sclerosis: a retrospective modeling and validation study. *Lancet Neurol* 21:246–257
16. Calabresi PA, Arnold DL, Sangurdekar D et al (2021) Temporal profile of serum neurofilament light in multiple sclerosis: implications for patient monitoring. *Mult Scler J* 27:1497–1505
  17. Coerver EME, Fung WH, de Beukelaar J et al (2024) Discontinuation of first-line disease-modifying therapy in patients with stable multiple sclerosis: the DOT-MS randomized clinical trial. *JAMA Neurol*. <https://doi.org/10.1001/jamaneurol.2024.4164>
  18. Bridel C, Leurs CE, van Lierop ZYGJ et al (2021) Serum neurofilament light association with progression in natalizumab-treated patients with relapsing-remitting multiple sclerosis. *Neurology* 97:E1898–E1905
  19. van Lierop ZYGJ, Verberk IMW, van Uffelen KWJ et al (2022) Pre-analytical stability of serum biomarkers for neurological disease: neurofilament-light, glial fibrillary acidic protein and contactin-1. *Clin Chem Lab Med* 60:842–850
  20. Wessels MH, Van Lierop ZY, Noteboom S et al (2023) Serum glial fibrillary acidic protein in natalizumab-treated relapsing-remitting multiple sclerosis: an alternative to neurofilament light. *Mult Scler J* 29:1229–1239
  21. Yaldizli O, Benkert P, Achtnichts L et al (2025) Personalized treatment decision algorithms for the clinical application of serum neurofilament light chain in multiple sclerosis: a modified Delphi Study. *Mult Scler*. <https://doi.org/10.1177/13524585251335466>
  22. Bose G, Healy BC, Saxena S et al (2023) increasing neurofilament and glial fibrillary acidic protein after treatment discontinuation predicts multiple sclerosis disease activity. *Neurol Neuroimmunol Neuroinflamm*. <https://doi.org/10.1212/NXI.0000000000200167>
  23. Kuhle J, Plavina T, Barro C et al (2020) Neurofilament light levels are associated with long-term outcomes in multiple sclerosis. *Mult Scler* 26:1691–1699
  24. Fox RJ, Cree BAC, de Sèze J et al (2024) Temporal relationship between serum neurofilament light chain and radiologic disease activity in patients with multiple sclerosis. *Neurology*. <https://doi.org/10.1212/WNL.0000000000209357>
  25. Freedman MS, Gnanapavan S, Booth RA et al (2024) Guidance for use of neurofilament light chain as a cerebrospinal fluid and blood biomarker in multiple sclerosis management. *EBioMedicine* 101:104970

## Authors and Affiliations

W. H. Fung<sup>1</sup> · M. H. J. Wessels<sup>1</sup>  · E. M. E. Coerver<sup>1</sup> · J. de Beukelaar<sup>2</sup> · W. H. Bouvy<sup>3</sup> · L. R. Canta<sup>4</sup> · O. H. H. Gerlach<sup>5</sup> · E. Hoitsma<sup>6</sup> · E. L. J. Hoogervorst<sup>7</sup> · B. A. de Jong<sup>1</sup> · N. F. Kalkers<sup>8</sup> · Z. L. E. van Kempen<sup>1</sup> · H. Lövenich<sup>9</sup> · C. E. P. van Munster<sup>10</sup> · B. W. van Oosten<sup>1</sup> · J. Smolders<sup>11</sup> · A. Vennegoor<sup>12</sup> · E. M. P. E. Zeinstra<sup>13</sup> · M. Barrantes-Cepas<sup>14</sup> · G. Kooij<sup>15</sup> · M. M. Schoonheim<sup>14</sup> · B. I. Lissenberg-Witte<sup>16</sup> · B. Moraal<sup>17</sup> · F. Barkhof<sup>17,18</sup> · B. M. J. Uitdehaag<sup>1</sup> · J. Mostert<sup>19</sup> · C. E. Teunissen<sup>20</sup> · J. Killestein<sup>1</sup> · E. M. M. Strijbis<sup>1</sup>

✉ M. H. J. Wessels  
m.wessels@amsterdamumc.nl

- <sup>1</sup> MS Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam UMC location VUmc, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands
- <sup>2</sup> Albert Schweitzer Hospital, Neurology, Dordrecht, The Netherlands
- <sup>3</sup> Diaconessenhuis, Neurology, Utrecht, The Netherlands
- <sup>4</sup> Catharina Hospital, Neurology, Eindhoven, The Netherlands
- <sup>5</sup> Zuyderland Medical Center, Neurology, Sittard-Geleen, The Netherlands, and School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands
- <sup>6</sup> Alrijne Hospital, Neurology, Leiden, The Netherlands
- <sup>7</sup> St. Antonius Hospital, Neurology, Utrecht, The Netherlands
- <sup>8</sup> OLVG, Neurology, Amsterdam, The Netherlands
- <sup>9</sup> St. Jans Gasthuis, Neurology, Weert, The Netherlands
- <sup>10</sup> Amphia Hospital, Neurology, Breda, The Netherlands
- <sup>11</sup> Erasmus Medical Center, MS center ErasMS, Neurology & Immunology, Rotterdam, The Netherlands
- <sup>12</sup> Flevoziekenhuis, Neurology, Almere, The Netherlands
- <sup>13</sup> Isala Hospital, Neurology, Meppel, The Netherlands

- <sup>14</sup> MS Center Amsterdam, Anatomy and Neurosciences, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands
- <sup>15</sup> MS Center Amsterdam, Molecular Cell Biology and Immunology, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands
- <sup>16</sup> Department of Epidemiology and Data Science, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands
- <sup>17</sup> Radiology & Nuclear Medicine, MS Center Amsterdam, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam UMC location VUmc, Amsterdam, The Netherlands
- <sup>18</sup> Queen Square Institute of Neurology and Centre for Medical Image Computing, University College, London, UK
- <sup>19</sup> Rijnstate Hospital, Neurology, Arnhem, The Netherlands
- <sup>20</sup> Neurochemistry Laboratory, Department of Clinical Chemistry, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Amsterdam, The Netherlands