Plasma glial fibrillary acidic protein and neurofilament light chain in relation to disability worsening in multiple sclerosis

Pauwels, Ayla; Van Schependom, Jeroen; Devolder, Lindsay; Van Remoortel, Ann; Nagels, Guy; Bjerke, Maria; D'hooghe, Marie B

Published in:
Multiple Sclerosis

DOI:
10.1177/13524585221094224

Publication date:
2022

License:
Unspecified

Document Version:
Accepted author manuscript

ABSTRACT

Background: Predicting disability worsening in multiple sclerosis (MS) remains an important challenge. Glial fibrillary acidic protein (GFAP) and neurofilament light chain (NfL) seem promising biomarkers. Studies investigating blood GFAP in relation to longitudinal outcome measures in MS are scarce.

Objective: To compare plasma-GFAP (p-GFAP) and plasma-NfL (p-NfL) levels in relation to sustained disability worsening.

Methods: We measured baseline p-GFAP and p-NfL in a prospective cohort of 115 individuals with MS and 30 matched controls, using Single Molecule Array (Simoa). Disability worsening was defined as an increase in ≥1 of 3 measures (Expanded Disability Status Scale, Timed 25-foot walk, 9-Hole Peg test), confirmed after ≥6 months and persistent upon data closure.

Results: In a multivariable Cox proportional-hazards model, p-GFAP was not significantly associated with sustained disability worsening after 4.40±0.82 years, while p-NfL (HR=1.046, P=0.001), EDSS (HR=1.24, P=0.039) and disease duration (HR=1.048, P=0.017) were. Area under the curve of ROC curves in relation to worsening was 0.61 for p-GFAP (P=0.031) and 0.63 for p-NfL (P=0.015). Kaplan-Meier curves showed similar patterns for both proteins.

Conclusion: p-NfL emerged as a significant explanatory variable for worsening in Cox regression analysis, p-GFAP did not. Both p-GFAP and p-NfL were related to worsening based on ROC curves.

INTRODUCTION

As of yet, no biomarker has been established that can accurately predict future disability in multiple sclerosis (MS). Glial fibrillary acidic protein (GFAP) and neurofilament light chain (NfL) have been suggested as candidate biomarkers for disability worsening in MS[1–4].

GFAP is the major cytoskeletal protein in astrocytes and is a marker of astrogliosis, a complex process of astrocytic changes upon CNS damage[5,6]. GFAP in cerebrospinal fluid (CSF) and blood shows promise in distinguishing MS phenotypes (CSF[4,7], blood[6,9]) and correlates with cross-sectional disease severity and MRI metrics (CSF[2,4,8,10], blood[6,8,11,12]). CSF GFAP seems associated with future disability[2–4]. Given the moderate to strong correlation between CSF and blood GFAP[8–10] and the less invasive nature of a blood draw versus lumbar puncture, the potential of blood GFAP as prognostic biomarker for MS disease course is worth investigating further.

NfL is an axonal cytoskeletal protein that is released in the CSF – and subsequently in the blood – upon neuroaxonal damage[1]. Similar to GFAP, the moderate to strong correlation of NfL in blood vs. CSF suggests that blood NfL might be used as a proxy for CSF measurements[1]. Multiple studies found that NfL levels in CSF and blood are higher in people with MS than in healthy controls (HC) and correlate with cross-sectional disability measures, relapse activity, and MRI metrics[1,13]. Studies on NfL in relation to long-term disability worsening show inconsistent results[14–19]. It has been suggested that combining NfL with other biomarkers might improve its performance[1].

The primary objectives of this study were to assess whether plasma-GFAP (p-GFAP) is associated with disability worsening in MS and to compare the association of p-GFAP and plasma-NfL (p-NfL) with (time to) disability worsening in a heterogeneous cohort of both progressive and relapsing MS patients. A secondary objective was to assess whether there is added value of combining p-GFAP and p-NfL in relation to disability worsening.

METHODS

Cohort:

In 2014, patients diagnosed with MS (according to the 2010 McDonald Criteria[20]) and a healthy control group (matched for age, sex, education) were recruited at the National MS Center Melsbroek (NMSC) and Brussels University Hospital (UZB) in Belgium for a prospective cohort study. For detailed in- and exclusion criteria, we refer to this study[21].

At baseline, participants underwent neurological assessments and had a blood sampling. Neurologic evaluation was repeated three months later and then during follow-up study visits one, two and four to five years (patients only) post-baseline. Additional patient data from in-between study visits was obtained from clinical follow-up records in the NMSC or UZB.

Clinical data:

During each study visit, neurological disability of MS participants was assessed with the Expanded Disability Status Scale (EDSS) for overall disability[22], Timed 25-Foot Walk (T25FW) for short-distance walking speed[23] and 9-Hole Peg Test (9-HPT) for manual dexterity[24]. Baseline MS Severity Score (MSSS) was determined[25]. Disability worsening was defined with the EDSS-Plus[26], a composite measure defined as worsening in ≥1 of 3 measures (EDSS[22], 9-HPT[24], T25FW[23]). Worsening needed to be sustained, i.e. confirmed after ≥6 months and persistent throughout the last study visit. The EDSS is a non-linear scale, with clinical relevance of observed changes depending on baseline values[22,27]: for patients with EDSS=0 at baseline, an increase of 1.5 points is needed to indicate worsening; for patients with a baseline EDSS of 1.0 to 5.5, a 1-point increase must be detected; for patients with an EDSS≥6 at baseline, an increase by 0.5 point suffices. For T25FW and 9-HPT, meaningful worsening is defined as a ≥20% score increase in scores.

Time to EDSS-Plus worsening was determined from the date of the baseline study visit to the date of initial increase in disability on the EDSS and/or T25FW and/or 9-HPT. Time to

worsening data was handled as missing if there were more than twelve months between the last stable assessment and the first measurement of increased disability.

Relapses were defined as new or worsening focal neurological symptoms lasting for more than 24 hours and usually followed by at least partial recovery over time[27]. Atypical symptoms or symptom worsening in the context of infection were not considered relapses. Relapses were assessed during study visits and clinical visits in-between and were allowed to be patient-reported.

Disease-modifying treatment (DMT) during the study period was categorized as follows: untreated (maximum six months treated), treated (maximum six months untreated), partially treated (more than six months treated but also more than six months untreated).

Blood biomarker analysis:

At baseline, a venous blood sampling was performed. EDTA tubes were centrifuged at 3500 rpm during 15 minutes at 3-5°C. Plasma supernatant was transferred to 1ml cryotubes and stored at -80°C. Biomarker analyses were performed blinded to clinical data by a board-certified laboratory technician. NfL and GFAP were determined using the Single Molecule Array (Simoa®) HD-1 platform (Simoa® NF-light™ Advantage Kit and GFAP* Discovery Kit, respectively; Quanterix, MA, USA) following manufacturer's instructions. All samples were within the measurement range of the assays. High- and low-quality control (QC) samples were analyzed in duplicates in the beginning, middle and end of each run for both p-NfL and p-GFAP. The mean intra-assay coefficients of variation (CVs) for p-NfL were 7.1% (QC low) and 2.7% (QC high) and 2.7% (QC low) and 3.5% (QC high) for p-GFAP, with no single replicate value exceeding a CV of 15%. The mean inter-assay CVs were 2.3% (QC low) and 6.7% (QC high) for p-NfL and 5.5% (QC low) and 2.3% (QC high) for p-GFAP.

Ethical compliance:

Study procedures are compliant with all relevant ethical regulations, aligning with the Declaration of Helsinki and Belgian privacy laws. Study procedures were approved by the medical ethics committee of the University of Brussels-Brussels University Hospital (approval number B.U.N. 143201317985) and of the National MS Center Melsbroek. Written informed consent was obtained from all participants.

Statistical analysis

All statistical analyses and graphical representations were performed in R[28] version 1.4.1106. For packages used, see supplementary material. Absence of normality for p-GFAP and p-NfL was confirmed using the Shapiro Wilk test. Because of non-normal distributions, non-parametric tests were applied in all further analyses. Spearman correlations were used for correlations between numerical variables. Where necessary, age correction was performed, by means of partial correlation or multivariable logistic regression. Group differences were tested with the Wilcoxon Rank-sum test, with effect size *r* calculated by dividing the *z* statistic by the square root of the sample size[29]. The Chi-square test was used for unranked categorical data, with effect size φ based on the square root of χ^2 divided by sample size[29].

Univariable and multivariable Cox proportional-hazards regression models were used to evaluate the association between the candidate biomarkers (in their numerical form) and time to EDSS-Plus worsening. For the multivariable model, the work method was as follows: first, possible explanatory variables of worsening were put in a multivariable Cox regression model. Following baseline variables were included: EDSS, disease duration, age, sex, onset phenotype, DMT. Annualized relapse rate (ARR) and DMT during study follow-up (treated/untreated/partially treated) were added. Then, stepwise Akaike Information Criterion (AIC) was used to choose the model with the best fit for our dataset (i.e. lowest AIC). The linearity assumption (martingale residuals and deviance residuals) and proportional hazard's assumption were checked and deemed acceptable.

Plasma-GFAP and p-NfL were divided into 'high' and 'low' based on receiver operating characteristic (ROC) curves, with the cut-off based on optimal sensitivity and specificity. The p-value for the ROC curves was determined based on the principle that the AUC is equivalent to the Mann-Whitney U-statistic testing significance of forecast event probabilities for the actually worsening versus non-worsening group, as explained by[30]. In the subgroup of patients with time to worsening data, we performed a Kaplan-Meier time to event analysis with the dichotomized values of p-GFAP and p-NfL.

RESULTS

1. Cohort description

1.1. Baseline cohort description

The cohort consisted of 115 patients with MS (87 relapsing-remitting (RR)MS, 28 primary progressive (PP)MS) and 30 healthy controls (HC). Patients were interferon-beta treated (N=30) or untreated (N=85, of whom N=54 were treatment naïve). Patients with RRMS were younger and had lower disability than PPMS patients (see table 1). Five RRMS patients experienced a relapse during baseline (onset ≤30 days before sampling). Baseline demographics and clinical characteristics are summarized in table 1.

1.2. Description of worsening cohort

The MS cohort was followed over a period of 4.40±0.82 years (range 3.19-5.43). As depicted in figure 1, forty-two patients showed sustained worsening, 61 did not. Twelve patients had undetermined outcome (reasons for unknown outcome: patients lost to follow-up (N=3), unconfirmed worsening (N=3), stable EDSS scores but lacking 9-HPT or T25FW data at baseline (N=2) or at the last time point (N=2), or suspected interrater variability interfering with EDSS assessment (N=2)). For 37/42 worsened patients, time to disability worsening was available (missing data for three RRMS and two PPMS patients) and was 1.96±1.98 years

(median±IQR). For all characteristics of the worsened and non-worsened group see table 2. DMTs at the last visit are specified in supplementary table 1.

2. p-GFAP and p-NfL

Values for p-GFAP and p-NfL for all participants are summarized in table 1. The candidate biomarkers correlated significantly with each other (Spearman correlation, N=145, r_s=0.44, P=$3.29 \cdot 10^{-8}$, supplementary figure 1).

2.1. p-GFAP in relation to baseline variables

Plasma-GFAP correlated with age (Spearman correlation, N=145, r_s=0.27, P=$9.87 \cdot 10^{-4}$). Levels were higher in female than in male participants in the whole cohort (Wilcoxon rank sum test, N=145, *r*=0.18, P=0.03) and in the RRMS subset (Wilcoxon rank sum test, N=87, *r*=0.26, P=0.02). Sex differences were not significant for the PPMS and HC subgroups (P=0.16 and P=0.37 resp.). Median p-GFAP levels did not significantly differ between patients and HCs (P=0.22). Plasma-GFAP levels were higher in patients with PPMS than in patients with RRMS (post-hoc Dunn's test: Z=2.41, adjusted P=0.03) and HCs (Z=2.46, adjusted P=0.04). This was no longer significant after correcting for age (P=0.07). There were no significant associations between p-GFAP and any of the baseline disability measures after correcting for age, nor were there associations with DMT status (interferon-treated vs. untreated) or disease duration.

2.2. Baseline p-NfL in relation to baseline variables

Plasma-NfL levels correlated moderately with age (Spearman correlation, N=145, r_s=0.46, P=$5.90 \cdot 10^{-9}$). There was no association with sex. Baseline p-NfL was higher in the MS group than in HCs (Wilcoxon rank sum test, *r*=0.19, P=0.02). NfL was higher in PPMS versus RRMS (*r*=0.37, P=$3.06 \cdot 10^{-4}$), which was no longer significant after correcting for age (P=0.41). NfL showed a correlation with all baseline disability measures, which persisted after correcting for age: EDSS (Spearman correlation, N=115, r_s=0.26, P=0.006) and MSSS (N=115, r_s=0.19, P=0.045), T25FW (N=115, r_s=0.29, P=$1.92 \cdot 10^{-3}$), 9-HPT dominant (N=103, r_s=0.20, P=0.047),

9-HPT non-dominant (N=103, r_s=0.26, P=0.009) scores. There was no association with DMT or disease duration.

3. Baseline p-GFAP and p-NfL in relation to disability worsening

Median levels of both p-GFAP and p-NfL appeared to be higher in patients with versus without EDSS-Plus worsening, which reached significance for p-NfL in the whole MS cohort (Wilcoxon rank sum test, N=103, *r*=0.21, P=0.03; table 2) and in the PPMS subset (N=25, *r*=40, P=0.04; figure 2B). Plasma-GFAP levels were not significantly higher in worsening MS patients in the whole cohort or in any subset (P=0.06 for the whole MS cohort, P=0.08 for RRMS, P=0.32 for PPMS; figure 2A, 1C).

We then evaluated the separate and combined association of p-GFAP and p-NfL (as numerical variables) with EDSS-Plus worsening in univariable and multivariable Cox regression. As shown in table 3, p-NfL, baseline EDSS and disease duration were associated with higher risk for EDSS-Plus worsening in univariable Cox regression models. In the final multivariable Cox regression model, p-NfL, baseline age, baseline EDSS and baseline disease duration were selected in the model with stepwise AIC, with p-NfL, EDSS and disease duration at baseline being significant risk factors for EDSS-Plus worsening (see table 3). p-GFAP was not retained in the model. These results did not change when excluding the five patients with a relapse during the baseline visit (data not shown).

To make our results comparable to other studies using cut-offs, we determined the optimal cut-off for p-GFAP and p-NfL in relation to EDSS-Plus worsening using ROC curves (figure 3). For p-GFAP, this cut-off was 79.19 ng/L, with a sensitivity of 69% and specificity of 54% (N=103, AUC=0.61 [95% confidence interval=0.50-0.72], P=0.030). For p-NfL, the cut-off was 12.19 ng/L with a sensitivity of 60% and specificity of 62% (N=103, AUC=0.63 [95% CI=0.52-0.73], P=0.015). Finally, we performed a Kaplan-Meier survival analysis in the subgroup with available time to worsening data. The analysis on high (≥79.19ng/L) vs. low

GFAP in relation to time to EDSS-Plus worsening did not reach significance (N=98, log-rank test P=0.054; figure 4). Patients with high p-NfL (≥12.19ng/L) showed a significantly shorter time to EDSS-Plus worsening than patients with low p-NfL (N=98, log-rank test P=0.045; figure 4), which lost significance when excluding N=5 patients with a relapse at baseline (N=93, log-rank test P=0.058).

DISCUSSION

For several years, NfL has been the main blood-based biomarker of interest in MS[1]. However, ongoing inflammatory disease activity may limit the ability of NfL to reflect neurodegenerative aspects underlying disability accumulation in MS[27,31]. Consequently, NfL could benefit from adding a biomarker with other pathophysiological properties[1]. CSF GFAP has been shown to be predictive for future disability in longitudinal studies[2–4], rendering blood GFAP an appealing candidate biomarker for disability worsening. Data on blood GFAP in relation to disability worsening in MS is scarce.

We compared p-GFAP and p-NfL in relation to (time to) disability worsening as defined with the EDSS-Plus, and investigated whether MS patients with higher p-GFAP have increased risk for disability worsening.

In the univariable Cox proportional-hazards model, the hazard ratio of p-GFAP as numerical variable was mildly and non-significantly increased in relation to EDSS-Plus worsening, whereas p-NfL was significantly associated with risk for worsening. The fact that p-NfL remained significant as a numerical variable in the model that controlled for baseline age, disease duration and disability, might imply that p-NfL is a stronger predictor for worsening than p-GFAP. ROC curve analysis of p-GFAP and p-NfL in relation to EDSS-Plus worsening revealed an AUC which significantly differed from 0.5 for both proteins, indicating that p-GFAP and p-NfL are able to distinguish between worsening and non-worsening patients. The patterns of the Kaplan-Meier curves were similar for both proteins, and visual appraisal suggested that higher levels were associated with a shorter time to disability worsening. The log-rank test was however not significant for GFAP and lost significance for NfL after excluding relapse patients, implying that the significance of these findings is very much dependent on the number of observations. While our cut-off for p-NfL is in line with results from a large study on p-NfL in relation to worsening in MS[15], it also lies within the normal

range for HCs. The overlap of levels between people with MS and HCs is a known pitfall of NfL as biomarker in MS[15]. Our cut-off for GFAP also lies within the normal HC range (see table 1).

A previous study in PPMS found no associations between serum GFAP and NfL and disability worsening after five years, but the clinical evaluation relied on patient-reported outcomes[32]. Consistent with previous cross-sectional studies[6,8,11], we found a moderate correlation between NfL and GFAP. GFAP was not significantly higher in MS patients than in HCs, in contrast to two previous studies[6,33]. While Högel et al.[6] did find higher GFAP in the whole MS cohort (RRMS and SPMS) in comparison with the HCs, levels were similar in RRMS and HCs, like in our study[6]. We found higher GFAP in PPMS than in RRMS, which lost significance after correcting for age. This was consistent with one study[8], but not with another[9]. Our RRMS cohort has a low annualized relapse rate and the PPMS cohort remained relatively stable during the study period, features that reflect the longer disease duration of our study participants. This might explain some of the inconsistencies with previous studies.

GFAP levels were higher in the female versus male participants, which seemed driven by the RRMS group. The recent longitudinal study on GFAP in PPMS patients also found higher GFAP in female participants[32], but others did not report this sex-related difference[6,8,11]. Stratifying by sex did not render significant associations between GFAP and time to worsening (data not shown).

Similar to two previous studies[9,33], we found no associations between GFAP and disability measures at baseline. Abdelhak et al.[8,11] did find correlations between GFAP and EDSS in progressive MS but not in RRMS[8]. Possibly, cross-sectional disease severity is better reflected by GFAP in progressive MS than in RRMS, and our PPMS subset might be too small to replicate these results. It remains to be elucidated whether p-GFAP is influenced by relapse

activity[3,4,7,35]. Our study comprised too few patients sampled during a relapse to make conclusions about this.

Limitations of our study are the lack of brain MRI data and missing data from twelve patients on EDSS-Plus worsening. While we cannot exclude a potential bias, an equal percentage of people with RRMS and PPMS had undetermined outcome, and to our estimation this group included both patients with and without actual worsening. For reasons of transparency, their GFAP and NfL values are shown in figure 2. Five patients with unknown time to worsening were excluded from the survival analyses, which could have affected the power of our analysis. To tackle the heterogeneous nature of the cohort, we used a composite measure better suited to assess worsening in a heterogeneous cohort of both relapsing and progressive patients than EDSS alone[31]. The analyses with 'high' and 'low' p-NfL and p-GFAP were based on a cut-off determined on our own dataset, so these results should be replicated in an independent sample. In conclusion, both p-GFAP and p-NfL relate to the occurrence of sustained clinical worsening in this cohort, with p-NfL being significantly associated with time to clinical worsening based on the Cox regression analysis. Further studies are needed to assess whether the combination of both biomarkers might prove to be useful. The clinical relevance of our findings remains to be determined in an independent larger cohort of patients with relapsing and progressive MS both early and late in the disease course.

Acknowledgements

We thank all study participants and all who provided logistic support for sample collection (health practitioners in the National MS Center, Annick Van Merhaegen-Wieleman and Tatjana Reynders in Brussels University Hospital).

Funding

This work was co-funded by the Fund D.V., managed by the King Baudouin Foundation (grant number 2016-C5812060-206012), the National MS Society and an FWO research grant (G038318N) awarded to MD and Prof. Jeroen Raes.

Author contributions

Statistical analyses were carried out by AP with guidance from JVS, LD, GN. Blood sampling and recruiting of the baseline cohort was carried out by AVR. Cleaning of metadata was done by LD and AP. The idea of combining baseline p-NfL and p-GFAP was suggested by MB. Blood sample analysis incl. quality control evaluation was planned and executed by MB. Clinical assessment of participants during study visits was performed by MD, AP and AVR. Patient data from in-between study visits was collected and quality checked by AP. Evaluation of sustained worsening for each patient was performed by AP. The manuscript was drafted by AP, with critical revision by JVS, LD, AVR, GN, MB, MD. All authors approved the final version for publication.

Conflicting interests

The authors declare that there is no conflict of interest.

REFERENCES

1. Bittner S, Oh J, Havrdová EK, et al. The potential of serum neurofilament as biomarker for multiple sclerosis. *Brain*. Epub ahead of print 28 June 2021. DOI: 10.1093/brain/awab241.
2. Axelsson M, Malmeström C, Nilsson S, et al. Glial fibrillary acidic protein: A potential biomarker for progression in multiple sclerosis. *J Neurol* 2011; 258: 882–888.
3. Martínez MAM, Olsson B, Bau L, et al. Glial and neuronal markers in cerebrospinal fluid predict progression in multiple sclerosis. *Mult Scler J* 2015; 21: 550–561.
4. Norgren N, Sundström P, Svenningsson A, et al. Neurofilament and glial fibrillary acidic protein in multiple sclerosis. *Neurology* 2004; 63: 1586–1590.
5. Sofroniew M V. Astrogliosis. *Cold Spring Harb Perspect Biol*; 7. Epub ahead of print 2015. DOI: 10.1101/cshperspect.a020420.
6. Högel H, Rissanen E, Barro C, et al. Serum glial fibrillary acidic protein correlates with multiple sclerosis disease severity. *Mult Scler J* 2020; 26: 210–219.
7. Mañé-Martínez MA, Olsson B, Bau L, et al. Glial and neuronal markers in cerebrospinal fluid in different types of multiple sclerosis. *J Neuroimmunol* 2016; 299: 112–117.
8. Abdelhak A, Huss A, Kassubek J, et al. Serum GFAP as a biomarker for disease severity in multiple sclerosis. *Sci Rep* 2018; 8: 14798.
9. Ayrignac X, Le Bars E, Duflos C, et al. Serum GFAP in multiple sclerosis: correlation with disease type and MRI markers of disease severity. *Sci Rep* 2020; 10: 10923.
10. Avsar T, Korkmaz D, Tütüncü M, et al. Protein biomarkers for multiple sclerosis: Semi-quantitative analysis of cerebrospinal fluid candidate protein biomarkers in different forms of multiple sclerosis. *Mult Scler J* 2012; 18: 1081–1091.
11. Abdelhak A, Hottenrott T, Morenas-Rodríguez E, et al. Glial Activation Markers in CSF and Serum From Patients With Primary Progressive Multiple Sclerosis: Potential of Serum GFAP as Disease Severity Marker? *Front Neurol* 2019; 10: 280.
12. Saraste M, Bezukladova S, Matilainen M, et al. Increased serum glial fibrillary acidic protein associates with microstructural white matter damage in multiple sclerosis: GFAP and DTI. *Mult Scler Relat Disord*; 50. Epub ahead of print 1 May 2021. DOI: 10.1016/j.msard.2021.102810.
13. Disanto G, Barro C, Benkert P, et al. Serum Neurofilament light: A biomarker of neuronal damage in multiple sclerosis. *Ann Neurol* 2017; 81: 857–870.
14. Thebault S, Abdoli M, Fereshtehnejad SM, et al. Serum neurofilament light chain predicts long term clinical outcomes in multiple sclerosis. *Sci Rep* 2020; 10: 10381.
15. Manouchehrinia A, Stridh P, Khademi M, et al. Plasma neurofilament light levels are associated with risk of disability in multiple sclerosis. *Neurology* 2020; 94: e2457–e2467.
16. Cantó E, Barro C, Zhao C, et al. Association between Serum Neurofilament Light Chain Levels and Long-term Disease Course among Patients with Multiple Sclerosis Followed up for 12 Years. *JAMA Neurol* 2019; 76: 1359–1366.
17. Anderson V, Bentley E, Loveless S, et al. Serum neurofilament-light concentration and real-world outcome in MS. *J Neurol Sci*; 417. Epub ahead of print 15 October 2020. DOI: 10.1016/j.jns.2020.117079.
18. Chitnis T, Gonzalez C, Healy BC, et al. Neurofilament light chain serum levels correlate with 10-year MRI outcomes in multiple sclerosis. *Ann Clin Transl Neurol* 2018; 5: 1478–1491.
19. Petzold A, Steenwijk MD, Eikelenboom JM, et al. Elevated CSF neurofilament

proteins predict brain atrophy: A 15-year follow-up study. *Mult Scler* 2016; 22: 1154–1162.
20. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Ann Neurol* 2011; 69: 292–302.
21. Reynders T, Devolder L, Valles-Colomer M, et al. Gut microbiome variation is associated to Multiple Sclerosis phenotypic subtypes. *Ann Clin Transl Neurol* 2020; 7: 406–419.
22. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*. Epub ahead of print 1983. DOI: 10.1212/WNL.33.11.1444.
23. Motl RW, Cohen JA, Benedict R, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. SAGE Publications Ltd.
24. Feys P, Lamers I, Francis G, et al. The nine-hole peg test as a manual dexterity performance measure for multiple sclerosis. *Multiple Sclerosis Journal* 2017; 23: 711–720.
25. Roxburgh RHSR, Seaman SR, Masterman T, et al. Multiple sclerosis severity score: Using disability and disease duration to rate disease severity. *Neurology* 2005; 64: 1144–1151.
26. Cadavid D, Cohen JA, Freedman MS, et al. The EDSS-Plus, an improved endpoint for disability progression in secondary progressive multiple sclerosis. *Mult Scler* 2017; 23: 94–105.
27. Cree BAC, Hollenbach JA, Bove R, et al. Silent progression in disease activity–free relapsing multiple sclerosis. *Ann Neurol* 2019; 85: 653–666.
28. R Development Core Team R. *R: A Language and Environment for Statistical Computing*. Epub ahead of print 2011. DOI: 10.1007/978-3-540-74686-7.
29. Fritz CO, Morris PE, Richler JJ. Effect size estimates: Current use, calculations, and interpretation. *J Exp Psychol Gen* 2012; 141: 2–18.
30. Mason SJ, Graham NE. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q J R Meteorol Soc* 2002; 128: 2145–2166.
31. Kappos L, Wolinsky JS, Giovannoni G, et al. Contribution of Relapse-Independent Progression vs Relapse-Associated Worsening to Overall Confirmed Disability Accumulation in Typical Relapsing Multiple Sclerosis in a Pooled Analysis of 2 Randomized Clinical Trials. *JAMA Neurol* 2020; 77: 1132–1140.
32. Giarraputo J, Giamberardino S, Arvai S, et al. Profiling serum neurofilament light chain and glial fibrillary acidic protein in primary progressive multiple sclerosis. *J Neuroimmunol*; 354. Epub ahead of print 15 May 2021. DOI: 10.1016/j.jneuroim.2021.577541.
33. Liu C, Lu Y, Wang J, et al. Serum neurofilament light chain and glial fibrillary acidic protein in AQP4-IgG-seropositive neuromyelitis optica spectrum disorders and multiple sclerosis: A cohort study. *J Neurochem*. Epub ahead of print 2021. DOI: 10.1111/jnc.15478.
34. Thion MS, Low D, Silvin A, et al. Microbiome Influences Prenatal and Adult Microglia in a Sex-Specific Manner. *Cell* 2018; 172: 500-516.e16.
35. Kassubek R, Gorges M, Schocke M, et al. GFAP in early multiple sclerosis: A biomarker for inflammation. *Neurosci Lett* 2017; 657: 166–170.