EBioMedicine 56 (2020) 102807

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Clinical implications of serum neurofilament in newly diagnosed MS patients: A longitudinal multicentre cohort study

Stefan Bittner[a,1], Falk Steffen[a,1], Timo Uphaus[a], Muthuraman Muthuraman[a], Vinzenz Fleischer[a], Anke Salmen[b,c], Felix Luessi[a], Achim Berthele[d], Luisa Klotz[e], Sven G. Meuth[e], Antonios Bayas[f], Friedemann Paul[g], Hans-Peter Hartung[h], Ralf Linker[i], Christoph Heesen[j], Martin Stangel[k], Brigitte Wildemann[l], Florian Then Bergh[m], Björn Tackenberg[n], Tania Kuempfel[o], Frank Weber[p,q], Uwe K. Zettl[r], Ulf Ziemann[s], Hayrettin Tumani[t,u], Sergiu Groppa[a], Mark Mühlau[d,v], Carsten Lukas[w], Bernhard Hemmer[d,v], Heinz Wiendl[e], Ralf Gold[b], Frauke Zipp[a,*], KKNMS consortium

[a] Department of Neurology, Focus Program Translational Neuroscience (FTN) and Immunotherapy (FZI), Rhine Main Neuroscience Network (rmn²), University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckstr. 1, Mainz 55131, Germany
[b] Department of Neurology, St. Josef-Hospital, Ruhr-University Bochum, Bochum, Germany
[c] Department of Neurology, Inselspital, Bern University Hospital and University of Bern, Bern, Switzerland
[d] Department of Neurology, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany
[e] Department of Neurology, University Hospital Münster, Westfälische-Wilhelms-University Münster, Münster, Germany
[f] Department of Neurology, Universitätsklinikum Augsburg, Augsburg, Germany
[g] NeuroCure Clinical Research Center and Experimental and Clinical Research Center, Charité, Universitätsmedizin Berlin and Max Delbrueck Center for Molecular Medicine, Berlin, Germany
[h] Department of Neurology, Medical Faculty, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany
[i] Department of Neurology, University of Regensburg, Regensburg, Germany
[j] Institute for Neuroimmunology and Multiple Sclerosis, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany
[k] Clinical Neuroimmunology and Neurochemistry, Department of Neurology, Hannover Medical School, Hannover, Germany
[l] Department of Neurology, University of Heidelberg, Heidelberg, Germany
[m] Department of Neurology, University of Leipzig, Leipzig, Germany
[n] Center of Neuroimmunology, Philipps-University Marburg, Marburg, Germany
[o] Institute of Clinical Neuroimmunology, Ludwig Maximilian University of Munich, Munich, Germany
[p] Max-Planck Institute of Psychiatry, Munich, Germany
[q] Neurological Clinic, Sana Kliniken des Landkreises Cham, Cham, Germany
[r] Department of Neurology, Neuroimmunological Section, University of Rostock, Rostock, Germany
[s] Department of Neurology, University of Tuebingen, Tuebingen, Germany
[t] Department of Neurology, University of Ulm, Ulm, Germany
[u] Clinic of Neurology Dietenbronn, Schwendi, Germany
[v] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
[w] Department of Radiology, St. Josef-Hospital, Ruhr-University Bochum, Bochum, Germany

ARTICLE INFO

Article History:
Received 20 February 2020
Revised 6 April 2020
Accepted 5 May 2020
Available online xxx

Keywords:
Neurofilament light chain
sNfL
Multiple sclerosis
Prediction
Biomarker

ABSTRACT

Background: We aim to evaluate serum neurofilament light chain (sNfL), indicating neuroaxonal damage, as a biomarker at diagnosis in a large cohort of early multiple sclerosis (MS) patients.
Methods: In a multicentre prospective longitudinal observational cohort, patients with newly diagnosed relapsing-remitting MS (RRMS) or clinically isolated syndrome (CIS) were recruited between August 2010 and November 2015 in 22 centers. Clinical parameters, MRI, and sNfL levels (measured by single molecule array) were assessed at baseline and up to four-year follow-up.
Findings: Of 814 patients, 54.7% (445) were diagnosed with RRMS and 45.3% (369) with CIS when applying 2010 McDonald criteria (RRMS[2010] and CIS[2010]). After reclassification of CIS[2010] patients with existing CSF analysis, according to 2017 criteria, sNfL levels were lower in CIS[2017] than RRMS[2017] patients (9.1 pg/ml, IQR 6.2–13.7 pg/ml, $n = 45$; 10.8 pg/ml, IQR 7.4–20.1 pg/ml, $n = 213$; $p = 0.036$). sNfL levels corre-

* Corresponding author: Stefan Bittner or Frauke Zipp, University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckstr. 1, Mainz, 55131, Germany. Phone: +49(0)6131-17-2805 or +49(0)6131-17-7156.
E-mail addresses: bittner@uni-mainz.de (S. Bittner), zipp@uni-mainz.de (F. Zipp).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ebiom.2020.102807

lated with number of T2 and Gd+ lesions at baseline and future clinical relapses. Patients receiving disease-modifying therapy (DMT) during the first four years had higher baseline sNfL levels than DMT-naïve patients (11.8 pg/ml, IQR 7.5-20.7 pg/ml, *n* = 726; 9.7 pg/ml, IQR 6.4–15.3 pg/ml, *n* = 88). Therapy escalation decisions within this period were reflected by longitudinal changes in sNfL levels.

Interpretation: Assessment of sNfL increases diagnostic accuracy, is associated with disease course prognosis and may, particularly when measured longitudinally, facilitate therapeutic decisions.

Funding: Supported the German Federal Ministry for Education and Research, the German Research Council, and Hertie-Stiftung.

1. Introduction

Neuroaxonal damage is the major underlying pathologic substrate of disability accumulation over time in patients with relapsing-remitting multiple sclerosis (RRMS). Patients with existing signs of neuroaxonal loss are at a high risk of developing secondary progressive and disabling disease courses [1]. Various studies suggest that tissue loss quantified by optical coherence tomography (OCT) [2,3] or MRI assessment [4] could be used to predict subsequent disability. However, both methods require either standardised longitudinal measurements or entail technical challenges and have therefore not yet been implemented in routine management. Serum neurofilament light chain (sNfL) has recently been proposed as a possible candidate for a reliable, easy-to-use biomarker of neuroaxonal damage [5]. Neurofilament proteins are part of the neuronal cytoskeleton and are elevated in various neurological diseases associated with neuronal damage including neurodegenerative diseases [6] and stroke [7], as well as MS [8]. Neurofilament light chain proteins are released after axonal injury into the cerebrospinal fluid (CSF) and to a lesser extent into the peripheral blood, where they can be measured by highly sensitive single molecule assays (SiMoA) [5]. sNfL has been shown to correlate with brain and spinal cord atrophy, as well as clinical relapses, T2 lesion load, and gadolinium-enhancing (Gd+) lesions in patients with RRMS [8–11]. However, to take the next step towards translation of sNfL into routine clinical use, there is a need for the investigation of large-scale real-world cohorts and longitudinal intra-individual samples.

Recent findings by us and others have indicated that sNfL might serve as a biomarker from very early stages of MS, namely in patients with clinically isolated syndrome [8] (CIS), pediatric MS patients [12], or even in presymptomatic stages of the disease [13]. This raises the possibility that sNfL could be included in diagnostic algorithms. The diagnosis of MS is based on McDonald criteria, which were first presented in 2001 and which underwent regular revisions in 2005, 2010 and 2017 as our understanding of this autoimmune CNS disease improved. The basic pillars of the McDonald criteria are identification of CNS lesions, using surrogate markers, and their dissemination in time (DIT) and space (DIS).

The criteria enable the diagnosis of RRMS in patients already at the first clinical presentation and to differentiate it from CIS, in which the chronicity of established MS is not yet proven, as is reflected by a lack of either DIT or DIS [14]. The definition of DIT has changed markedly in the current 2017 McDonald revision by additionally including a) symptomatic Gd+ lesions [15] and b) the presence of oligoclonal bands (OCB) in the CSF as a substitute for clinical or imaging evidence of DIT. The intention was to define chronicity earlier in the disease course allowing an earlier initiation of disease-modifying therapy (DMT), which should improve long-term prognosis. Retrospective analysis of the 2017 McDonald criteria in different cohorts [16–18] has demonstrated that the current revision allows an earlier diagnosis of RRMS compared to the 2010 criteria, in particular through the use of OCB to fulfil the DIT criterion. Additional markers are warranted to select patients at high risk of developing future disability (beyond making the diagnosis of MS *per se*) in order to initiate early and effective treatment of RRMS patients.

We here address, in a multicentre approach in 814 patients from the German National MS cohort with newly diagnosed CIS/RRMS, the potential clinical implications of assessing sNfL in a real world setting for diagnosis, prognosis, and therapeutic decisions.

Research in context

Evidence before this study

Serum neurofilament light chain protein (sNfL) levels have been shown to correlate with neuroaxonal damage in multiple sclerosis (MS) and various other neurological disorders. We used the terms “neurofilament light chain”, “NfL”, “multiple sclerosis”, “MS”, “prediction”, and “McDonald criteria” in PubMed to find publications from any date up to February 14, 2020. Several publications support an earlier diagnosis of relapsing-remitting MS (RRMS) with the 2017 McDonald criteria compared to the 2010 criteria based on magnetic resonance imaging (MRI) or clinical experience. There is strong evidence that sNfL levels correlate with specific clinical and MRI parameters that are commonly used in practice to monitor disease progression and treatment response. However, we found no report that uses sNfL to investigate the value of the 2017 McDonald criteria in order to identify patients at risk of disease progression and its impact on treatment decision and only a limited number of small pilot studies about the value of sNfL in very early MS patients.

Added value of this study

We classified patients as clinically isolated syndrome (CIS) or early RRMS according to 2010 and 2017 McDonald criteria and linked this with sNfL, a peripheral blood biomarker capable of detecting neuroaxonal damage. Thereby, we showed in a prospective multicentre study that the revised McDonald criteria 2017 are superior to detect patients at risk of neuroaxonal damage. Furthermore, we investigate in detail the relation of sNfL and disease course, as well as treatment decisions within the first four years after diagnosis. The observations are of additional value since they have been made in a young and initially untreated cohort at disease onset. Until now, this prognostically relevant group has been underrepresented in robust publications with high numbers of participants.

Implications of all the available evidence

In a prospective, multicentre cohort of 814 patients with CIS and early RRMS at disease onset, assessment of sNfL increased diagnostic accuracy and facilitated prognosis of the disease course over the next four years. Especially when it was measured longitudinally, sNfL reflected therapeutic decisions and may facilitate early therapeutic stratification of patients. Longitudinal measurement of sNfL rather than absolute cut-off values are recommended for clinical decision-making.

2. Materials and methods

2.1. Mainz cohort

To assess the impact of 2010 versus 2017 McDonald criteria on patients with newly diagnosed CIS and different levels of neuroaxonal damage, we first performed a retrospective cross-sectional single centre pilot study. Blood specimens of MS patients were collected and processed at the University Medical Center Mainz as described below. Datasets were available for 61 patients who had sNfL measurements, MRI measurements, and CSF analysis at baseline. These patients were initially classified according to 2010 McDonald criteria, and patients with an initial diagnosis of CIS were reclassified according to 2017 McDonald criteria.

2.2. German National MS cohort

The German National MS (NationMS) cohort is a multicentre prospective longitudinal observational study comprising (a) detailed assessment of patients with first diagnosis of MS or CIS according to 2005 McDonald criteria and (b) yearly assessment with a standardised protocol across 22 centres in Germany. All centres belong to the German Competence Network Multiple Sclerosis (KKNMS). The study was approved by the ethics committee of Ruhr-University Bochum (Registration no. 3714-10), and subsequently, by all local ethics committees of the participating centres. All patients provided written informed consent. Inclusion and exclusion criteria as well as assessment plans have been described previously [19]. All patients (n = 1,124) were included at least 30 days after relapse, but prior to DMT initiation. Thereafter, DMT was initiated in a "real world" setting by each centre. Complete datasets were available for 814 patients who had sNfL measurements and MRI at baseline and were followed up at least two years. For an additional 598 patients, clinical parameters were available after four years of follow-up. To exclude a selection bias, we analyzed baseline and clinical characteristics of both cohorts (814 vs 1124 patients in the whole NationMS cohort) for age, sex, disease duration, first clinical manifestation, and extended disability status scale (EDSS) and found no major differences for all parameters arguing against a selection bias. In our cohort the median age at inclusion was 33 years (IQR 26-41 years) compared to 32.4 years (26.6–41.0 years) in the total cohort. We reported a baseline EDSS median of 1·5 (IQR 1.0-2.0) which is in line with the median reported in the entire cohort ([19] and Table 1). For our study, patients were classified according to 2010 McDonald criteria, and patients with an initial diagnosis of CIS were reclassified according to 2017 McDonald criteria.

2.3. sNfL measurements

To ensure a high degree of standardisation, blood withdrawal was performed at the same day of MRI investigation but before the application of contrast medium using a standard protocol in all centres. Blood was collected in 10 ml Serum-Vacutainer®-tubes (Becton Dickinson, USA); samples were spun at 1300 g at room temperature for 15 min within 2 h after sampling. Directly after centrifugation, the serum was evenly transferred (300–600 μl/tube) in 1·1 ml polypropylene tubes and locally stored at −80 °C. Serum samples from all centres were then sent on dry ice to the KKNMS biobank and centrally stored at −80 °C.

In a previous study, several days of processing did not significantly affect NfL levels in plasma, indicating stability of the protein and a robust assay procedure [20]. For this project, serum samples of our included patients were sent on dry ice from the central biobank in Munich to Mainz. Here, measurements from the multicentre cohort were performed in one single centre at one machine with a standardised protocol and a single batch. sNfL was measured in several rounds by SiMoA HD-1 (Quanterix, USA) using the NF-Light Advantage Kits (Quanterix) from the same batch according to manufacturer's instructions. Resorufin-β-D-galactopyranoside (RGP) was incubated at 33 °C for 60 min prior to running the assay.

Samples were measured in duplicates. The coefficient of variation (CV, as a percentage) of each sample was obtained by dividing the

Table 1
Clinical and demographic data of CIS/early RRMS patients according to 2010 McDonald criteria included in this study at baseline.

Variable		CIS	RRMS	p-value
n		369	445	
		Median (IQR)		
sNfL (pg/mL)		10.1 (6.9–17.1)	12.5 (8.0–22.8)	<0.0005[a]
Age (years)		33 (26–41)	31 (26–40)	0.272[a]
EDSS		1.5 (1–2)	1.5 (1–2)	0.010[a]
		Mean ± SD: 1.3 ± 1.0	1.5 ± 1.0	
Disease duration (months)		2 (1–2)	2 (1–3)	0.141[a]
		n (%)		p-value
Sex	male	113 (30.6%)	146 (32.8%)	0.505[b]
	female	256 (69.4%)	299 (67.2%)	
OCB	neg.	23 (6.2%)	23 (8.9%)	0.593[b]
	pos.	188 (50.9%)	222 (86.0%)	
	unknown	158 (42.8%)	13 (5.0%)	
T2-lesion count	1–8	110 (30.1%)	132 (29.8%)	0.936[b]
	> 8	256 (69.9%)	311 (70.2%)	
	unknown	3 (0.8%)	2 (0.4%)	
GD-enhancement	no	269 (75.6%)	218 (50.3%)	<0.0005[b]
	yes	85 (23.9%)	213 (49.2%)	
	unknown	2 (0.6%)	2 (0.5%)	
Ring-enhancement	no	321 (93.9%)	364 (89.0%)	0.019[b]
	yes	21 (6.1%)	45 (11.0%)	
	unknown	27 (7.3%)	36 (8.1%)	
Treatment	no treatment	369 (100%)	445 (100%)	

IQR: interquartile range; CIS: clinically isolated syndrome; RRMS: relapsing-remitting multiple sclerosis; sNfL: serum neurofilament light chain; EDSS: Expanded Disability Status Scale; OCB: oligoclonal bands; SD: standard deviation.
[a]Mann–Whitney-U tests were conducted to compare group differences.
[b]Distributions were compared using chi-square tests of homogeneity.

standard deviation of both replicates by the mean of both replicates multiplied by 100. Since the range of sNfL concentrations in serum is smaller than in CSF, some samples with a sample CV above 20% (or missing replicate result) were measured twice, as in previous publications [8,9]. Finally, the mean intra-assay CV of 6.2% was obtained by averaging all individual sample CVs. The two same low and high controls, consisting of recombinant human NfL antigen, were run in duplicates with each sample run to monitor plate-to-plate variation. The mean concentration over all runs was 4.4 pg/ml for the low control and 141.5 pg/ml for the high control. We obtained inter-assay CVs of 6.0% and 13.2% for the low and high control, respectively. sNfL measurements were performed in a blinded fashion without information about clinical data.

2.4. Multiple Sclerosis Functional Composite (MSFC)

The MSFC is a score composed of three objective quantitative neurological tests for assessing arm, leg, and cognitive function and was developed to improve the measurement of clinical outcome in extension to the EDSS [21]. Subtests of the MSFC in the NationMS cohort have recently been published to investigate changes in cognitive impairment from baseline up to 12 months [22]. We here used our baseline cohort as a reference population for the z-standardisation instead of normative control cohorts in the previous publication, which leads to different z-scores without affecting the ratio of the amount of change to the standard deviation of the change. A detailed description of the administration and calculation of the MSFC were already published [23]. Z-scores were obtained by standardising all subtest scores to the baseline results of all patients included in this study. Finally, the MSFC score was calculated using the following formula: MSFC score = (Z-score $_{TW}$ + Z-score $_{9\text{-HPT}}$ + Z-score $_{PASAT}$)/3. (TW = Timed 25-Foot Walk; 9-HPT = Nine-Hole Peg Test; PASAT = Paced Auditory Serial Addition Test).

2.5. Statistics

Statistical analyses were performed with SPSS version 23 (IBM Corp., USA), MATLAB R2018a (MathWorks, USA) and RStudio version 1.1.456 (RStudio Inc., USA). The normal distribution of data was tested using the Kolmogorov–Smirnov and Shapiro–Wilk tests. We used a Mann–Whitney test or Kruskal–Wallis-Test with adjusted p values by Bonferroni correction, as appropriate. Effect size after applying a Mann–Whitney-U test was estimated using the formula ($r^2 = \eta^2 = \frac{Z}{\sqrt{n}}$) where Z is the standardised value for the U-value, r the correlation coefficient, and r^2 or η^2 indicate the percentage of variance in the dependent variable that can be explained by the independent variable when multiplied by 100%. Non-parametric correlation was determined by Spearman's rank correlation coefficient and partial non-parametric correlation when considering age as a covariate. Age as cofounding factor needs to be taken into account in older patients as sNfL seems to considerably increase in particular above the age of 60 years with rather stable values in younger patients [24]. In agreement, in our cohort of mainly young patients (median age 32 years, percentage of patients > 60 years old: 0.7%), we found no significant correlation between age and sNfL values (r = −0.044, p = 0.21, Supplementary Fig. 1A+B) and hence no age correction was necessary. However, all analyses were additionally performed with age as a covariate using one-way or two-way mixed ANCOVA where appropriate and can be found in Supplementary Table 1. Within-subject factors over time were analyzed by mixed linear models or two-way mixed ANOVA after log-transformation of sNfL values. This was followed by one-way ANOVA with Tukey post-hoc test for multiple comparison to calculate simple main effects. Delta sNfL values (two-year follow-up minus baseline) were reflected and set to a minimum of 1 followed by non-negative transformation of the resulting values. A Chi-square test of homogeneity was applied to investigate differences in proportions. When performed, post hoc analysis involved pairwise comparison using multiple z-test of two proportion with a Bonferroni correction. Boxplots are shown with the median represented by a horizontal line. Boxes extend from the 25th to 75th percentile. The upper whiskers expand from the 75th percentile to the highest value that is smaller than or equal to the interquartile range (IQR) multiplied by 1.5 and added to the 75th percentile. The lower whiskers extend from the 25th percentile to the smallest value that is higher or equal to the IQR multiplied by 1.5 and subtracted from the 25th percentile. For better visualisation, scatterplots were graphically modified using the syntax command *dodge* (Fig. 1A) or *jitter* (all other figures) in SPSS. All statistical analyses were performed using the original data without modifications. P values < 0.05 were considered statistically significant.

2.6. Bayesian analyses

To deal with imbalanced sample sizes for data with a non-Gaussian distribution (Fig. 2C and D), we used the Bayesian posterior distribution analyses as additional validation of significant group differences already determined by Mann–Whitney-U tests. This analysis provides complete distribution of credible values for group means and their differences [25]. Specifically, we tested for sNfL markers based on the two groups with and without taking age as a covariate for the capability of credible separation.

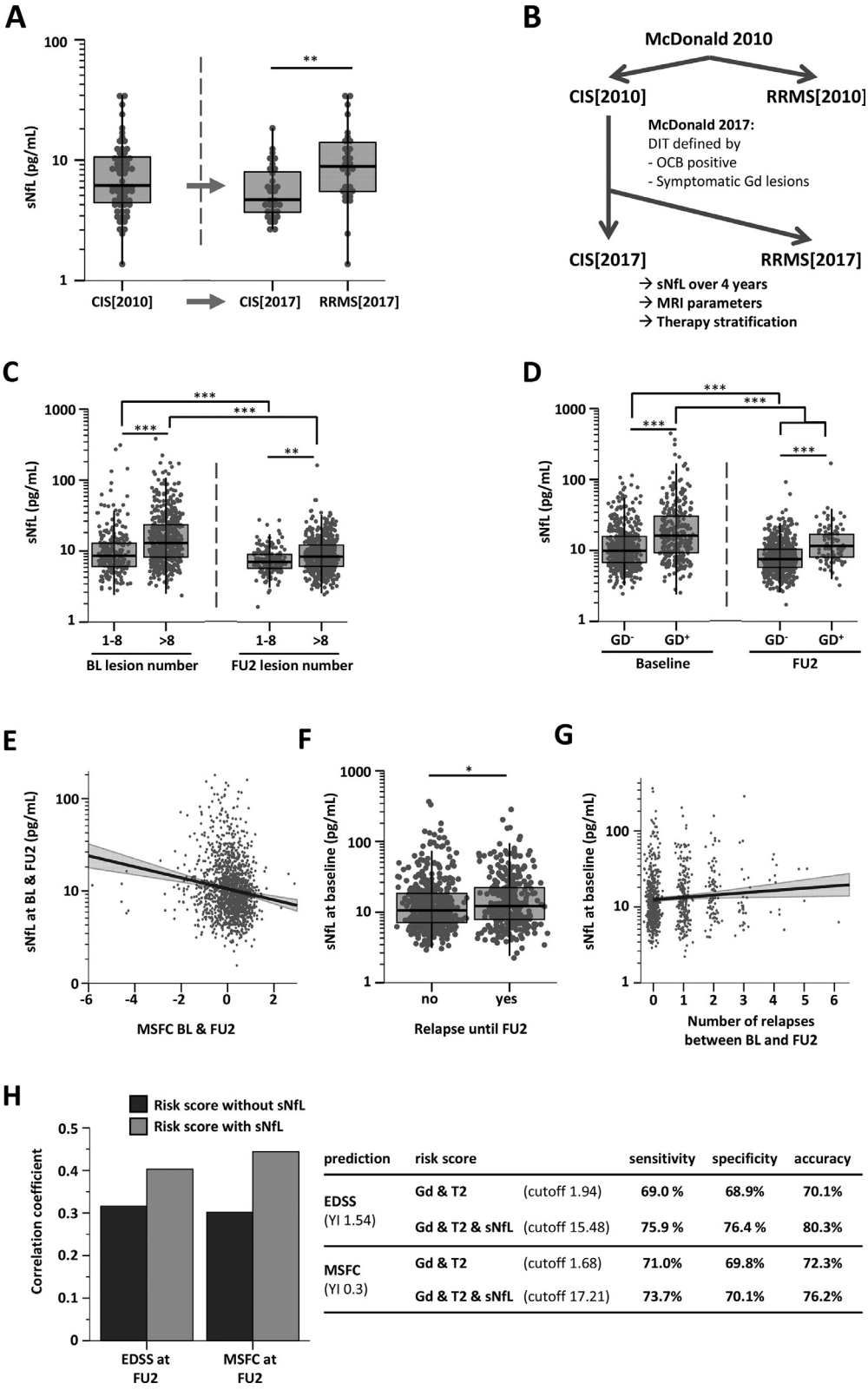
2.7. Composite score analyses

Composite scores were calculated in a two-step procedure. First, we performed a cluster analysis by grouping all possible combinations (always a pair) for the variables that had an area under the curve (AUC) > 0.5. Second, from the significant ($p < 0.05$) clusters corrected for multiple comparisons using Bonferroni corrections, we estimated a composite score using the partial least squares method (PLS) to assign the weights for each combination [26].

See also Supplementary Methods for more detailed methods.

3. Results

The overall aim of the study was to assess potential clinical implications of measuring sNfL for diagnostic accuracy, prognosis, and therapeutic decisions in early MS patients. To evaluate the impact of 2010 versus 2017 McDonald criteria on patients with newly diagnosed CIS and different levels of neuroaxonal damage, we first performed a pilot study. Patients diagnosed with CIS according to 2010 McDonald criteria (CIS[2010]) were reassessed and classified either as CIS or RRMS based on the 2017 criteria (CIS[2017], RRMS[2017]). Interestingly, in this single-center cohort, sNfL levels were higher in RRMS (8.9 pg/ml, IQR 5.5–14.3 pg/ml, n = 30) than in CIS patients according to the new criteria (4.7 pg/ml, IQR 3.6–8.5, p = 0·001, Fig. 1A). Based on this promising data, we designed a multicentre approach (Fig. 1B) where 814 patients were included at baseline and up to four-year follow-up (for baseline characteristics see Table 1). At baseline and year two, patients with ≥9 cranial T2 lesions according to Barkhof criteria [27] had significantly higher sNfL levels (baseline: 13.0 pg/ml, IQR 8.2–23.5 pg/ml, n = 567; two year follow-up: 8.4 pg/ml, 6.1–12.2 pg/ml, n = 573) than those with 1-8 T2 lesions (baseline: 8.6 pg/ml, IQR 6.1–12.9 pg/ml, n = 242; $p < 0.0005$; two-year follow-up: 7.0 pg/ml, 5.7–9.0 pg/ml, n = 161, p = 0.001, Fig. 1C). Comparable findings were obtained for sNfL and Gd+ lesions as Gd+ lesions correlated with high sNfL levels in all patients (Fig. 1D). Furthermore, we found a weak correlation between sNfL and EDSS values (r = 0.13, $p <$ 0.0005), between sNfL levels and ring enhancing lesions (Supplementary Fig. 2A+B), and an inverse correlation between sNfL levels and MSFC score at baseline and two-year follow-up (r = -0.170, $p <$ 0.0005, Fig. 1E). Patients who suffered from at least one relapse in the following two years had significantly higher levels at baseline



prediction	risk score		sensitivity	specificity	accuracy
EDSS (YI 1.54)	Gd & T2	(cutoff 1.94)	69.0 %	68.9%	70.1%
	Gd & T2 & sNfL	(cutoff 15.48)	75.9 %	76.4 %	80.3%
MSFC (YI 0.3)	Gd & T2	(cutoff 1.68)	71.0%	69.8%	72.3%
	Gd & T2 & sNfL	(cutoff 17.21)	73.7%	70.1%	76.2%

Fig. 1. sNfL at time point of diagnosis correlates with baseline MRI parameters in a multicentre cohort and predicts clinical activity within the next two years. **A)** In a single-centre pilot study, patients with CIS according to 2010 McDonald criteria (n = 61; CIS[2010]) were reclassified based on the 2017 version of the McDonald criteria. sNfL levels were significantly higher in patients switching to RRMS (8.9 pg/ml, IQR 5.5-14.3 pg/ml, n = 30; RRMS[2017]) compared to patients remaining CIS (4.7 pg/ml, IQR 3.6-8.5 pg/ml, n = 31; CIS [2017]; p = 0.001 determined by Mann−Whitney-U test). **B)** Study design of a multicentre approach to assess reclassification of CIS[2010] patients according to 2017 McDonald criteria. **C)** 1543 MRIs from 809 patients were assessed for sNfL levels comparing MRIs with > 8 T2 lesions (baseline: n = 567; two-year follow-up: n = 573) and 1-8 T2 lesions (baseline: n = 242; two-year follow-up: n = 161). sNfL was higher in patients with > 8 than 1-8 T2 lesions at baseline (13.0 pg/ml, IQR 8.2-23.5 pg/ml; 8.6 pg/ml, IQR 6.1-12.9 pg/ml, $p < 0.0005$) and two-year follow-up (8.4 pg/ml, IQR 6.1-12.2 pg/ml; 7.0 pg/ml, IQR 5.7-9.0 pg/ml, p = 0.001). Both lesion groups showed a significant decrease of sNfL concentration after two years. **D)** Patients with Gd+ lesions at the time of serum sampling (n = 298; two-year follow-up: n = 94) had significantly higher sNfL levels than patients without Gd+ lesions (n = 487; two-year follow-up: n = 606). **E)** Baseline and two-year follow-up sNfL levels were negatively correlated with corresponding MSFC values (r = -0.170, $p < 0.0005$). **F)** sNfL

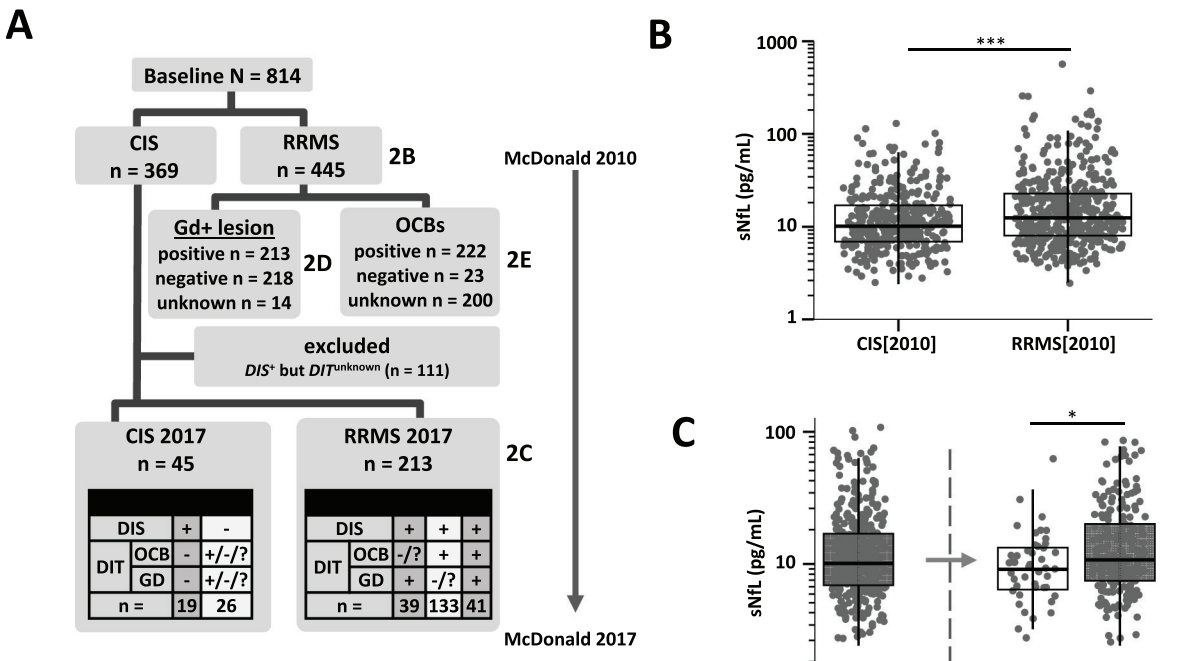
compared to patients experiencing no further relapses (12.2 pg/ml, IQR: 7.9–22.2 pg/ml, n = 337; no relapse: 10.6 pg/ml, IQR: 7.9–22.2 pg/ml, n = 477; p = 0.015, Fig. 1F). In agreement with this, we found a correlation between baseline sNfL levels and the absolute number of relapses in this two-year period (r = 0.092, p = 0.008, Fig. 1G). In order to further evaluate the predictive value of baseline sNfL values after diagnosis for future risk stratification, we developed a risk score using support vector machine (SVM) algorithms (see Supplementary Material and Methods). Using the two MRI parameters i) presence or absence of Gd+ lesions and ii) 1-8 T2 lesions or more, a composite risk score was assigned to each individual patient at baseline. The predictive capacity of the risk score for the two outcome parameters EDSS and MSFC in a two-year follow-up was calculated by SVM analysis. Correlation coefficients drastically increased by including the sNfL baseline values in the composite risk score (EDSS: 0.32 to 0.40, MSFC: 0.30 to 0.44, Fig. 1H), underlining that evaluation of both MRI parameters and sNfL had an added-value compared to only using one single approach.

Out of the 814 patients at disease onset, 45.3% (n = 369) had a diagnosis of CIS[2010] and 54.7% (n = 445) had a diagnosis of RRMS [2010] (Fig. 2A). sNfL levels were higher in RRMS[2010] (12.5 pg/ml, IQR 8.0–22.8 pg/ml, n = 445) than in CIS[2010] (10.1 pg/ml, IQR 6.9-17.1 pg/ml, n = 369, $p < 0.0005$, Fig. 2B) despite similar baseline characteristics such as disease duration (Table 1). When applying the 2017 McDonald criteria to the same patients, both the presence of OCB and the assessment of symptomatic Gd+ lesions could change the classification of CIS[2010] to RRMS[2017]. We excluded 111 patients due to missing information regarding DIT. To prevent a selection bias, subgroup analyses were performed showing no significant differences (see Supplementary Table 2). Upon reclassification of CIS[2010] patients with existing CSF analysis, according to 2017 criteria, only 17.4% (45/258) remained CIS[2017] whilst 82.6% (213/258) were reclassified as RRMS[2017]. Importantly, patients who were reclassified from CIS[2010] to RRMS[2017] had elevated sNfL levels (10.8 pg/ml, IQR 7.4–20.1 pg/ml, n = 213) compared to CIS[2010] patients remaining CIS[2017] (9.1 pg/ml, IQR 6.2-13.7 pg/ml, n = 45, p = 0.036, Fig. 2C). Taking into account the imbalanced sample sizes, this analysis was additionally confirmed by Bayesian analysis, which resulted in a discrimination accuracy of 93.0%. These findings demonstrate that application of 2017 McDonald criteria to CIS patients results in diagnosis of RRMS compared to a diagnosis of CIS when applying the previous McDonald criteria in patients with increased neuroaxonal damage. Additionally, we evaluated whether the presence or absence of OCB and Gd+ in RRMS[2010] patients separates patients with high and low sNfL levels. Both OCB-positive RRMS [2010] patients (13.1 pg/ml, IQR 8.0–24.3 pg/ml, n = 222) and RRMS [2010] patients with Gd+ lesions (16.4 pg/ml, IQR 9.3–31.3 pg/ml, n = 213) had significantly higher sNfL levels than OCB-negative (10.1 pg/ml, IQR 6.7–15.8 pg/ml, n = 23, p = 0.035, accuracy of discrimination = 96.4%) and Gd-negative (10.2 pg/ml, IQR 7.2–16.1 pg/ml, n = 218, $p < 0.0005$) patients (Fig. 2D+E).

To unravel whether inclusion of sNfL in the McDonald diagnostic criteria algorithm would increase the discrimination accuracy between patients with CIS and RRMS, receiver operating characteristic (ROC) analysis was performed in order to reclassify CIS[2010] as CIS[2017] or RRMS[2017]. Inclusions of the 90th percentile of sNfL (31.2 pg/ml) led to a significantly increased (p = 0.035) area under the curve (AUC = 0.84, CI 0.79– 0.89, $p < 0.0005$) compared to OCB and/or Gd+ (AUC = 0.76, CI 0.70–0.83, $p < 0.0005$, Fig. 2F). We confirmed these linear classification data by predictive analysis using a machine learning algorithm SVM which is a non-linear classifier. Importantly, the prediction accuracy of OCB and/or Gd+ (sensitivity: 72%, specificity: 76%, accuracy: 79%) were again further increased by including the 90th percentile of sNfL in addition to the above two variables (sensitivity: 73%; specificity: 79%, accuracy: 84%; Fig. 2G, for more data on 50th to 90th percentile see Supplementary Table 3). These findings point towards a potential value of especially high sNfL levels (>31 pg/ml) at time of first demyelinating event as indicators of ongoing chronic CNS neuroinflammation and may be considered for inclusion in a future refinement of the McDonald criteria.

The changes in the 2017 McDonald criteria are intended to allow earlier diagnosis of RRMS and thus to facilitate initiation of DMT as early as possible in these patients [28]. Therefore, we assessed whether sNfL levels can predict later initiation of DMT and whether treatment would influence sNfL levels after two years of follow-up. The percentage of patients under therapy at two-year follow-up was comparable between CIS[2010] (76% (279/369)) and RRMS[2010] (81% (359/445)) patients (Fig. 3A). Reclassification of CIS[2010] patients according to 2017 criteria had no impact on whether DMT were administered or not (Fig. 3A). At baseline, all patients were treatment naïve. Indeed, patients without DMT initiation during the first two years showed lower sNfL levels at baseline (9.5 pg/ml, IQR 6.4–14.1 pg/ml, n = 87) than patients with at least one (transient) DMT during the observation period (11.8 pg/ml, IQR 7.5–20.9 pg/ml, n = 727, p = 0.002, Fig. 3B). We next grouped patients into four classes based on the type of DMT they were receiving at two-year follow-up ("no DMT": n = 176; "basic": interferons and glatirameracetate, n = 392; "moderate": teriflunomide and dimethylfumarate, n = 134; and "high": natalizumab, rituximab, fingolimod, ocrelizumab, daclizumab, alemtuzumab, mitoxantrone, n = 107). Five patients were excluded from DMT analysis due to unknown treatment. Of note, sNfL baseline levels significantly correlated with the established treatment group at two-year follow-up (r=0.223, p<0.0005, Fig. 3C). While all treatment groups had comparable sNfL levels after two years (no DMT: 8.4 pg/ml, IQR 6.2–11.6 pg/ml; "basic": 7.7 pg/ml, IQR 5.8–11.4 pg/ml; "moderate": 7.5 pg/ml, IQR 5.6–11.0 pg/ml; "high": 8.2 pg/ml, IQR 6.4–11.6 pg/ml; all $p > 0.05$), patients on "high" therapies had significantly elevated baseline levels ("high": 21.0 pg/ml, IQR 12.0–45.3 pg/ml) compared to other groups (no DMT: 10.0 pg/ml, 6.6–18.3 pg/ml, $p < 0.0005$; "basic": 10.5 pg/ml, 7.1–17.1 pg/ml, $p < 0.0005$; "moderate": 12.0 pg/ml, 7.4–22.0 pg/ml, $p < 0.0005$; Fig. 3D). This indicates that high sNfL levels at disease manifestation correlate with real world therapy decisions since sNfL levels were not known to treating physicians at the time point of treatment initiation. After adjustment for the baseline sNfL concentrations, comparing the delta sNfL values (two-year follow-up minus baseline) of the different treatment groups between each other further showed a massive sNfL decline in the "high" treatment group (median −11.3 pg/ml, IQR −37.3 to −2.9 pg/ml) that was significantly higher than in the "basic" group (−2.2 pg/ml, −7.6 to 0.8 pg/ml, p = 0.001) and no DMT group (−0.8 pg/ml, −6.5 to 1.6 pg/ml, p = 0.001, Fig. 3E). This might be due to resolving initial inflammation (regression to the mean phenomenon), as all patients were included at least 30 days after relapse, but prior to DMT initiation, and acute inflammatory

levels at baseline were significantly higher in patients suffering at least one relapse up to FU2 (median: 12.2 pg/ml, IQR: 7.9-22.2 pg/ml, n = 337) than patients without relapse (median: 10.6 pg/ml, IQR: 7.9-22.2 pg/ml, n = 477, p = 0.015). **G)** Baseline sNfL levels correlated with the number of relapses between baseline and FU2 (r = 0.092, p = 0.008). **H)** A baseline risk score consisting of either presence of Gd+ lesions and T2 lesions (1-8 lesions or more than 8) alone (EDSS cutoff point = 1.94; MSFC cutoff point = 1.68) or in addition to sNfL (EDSS cutoff point = 15.48; MSFC cutoff point = 17.21) was determined for risk stratification at study initiation. The predictive power of the model for the two outcome parameters EDSS (YI = 1.54) and MSFC (YI = 0.3) and at FU2 was assessed by SVM algorithm. Correlation coefficients were increased (EDSS: 0.32 to 0.40, MSFC: 0.30 to 0.44) by the inclusion of baseline sNfL levels in the risk score computation. 10-fold cross validation was performed for the correlation and accuracy parameters. Detailed information on the model specifications are depicted in the table. Group differences were analyzed by mixed linear model procedure or Mann-Whitney-U test. Correlation analysis was performed by Spearman's rank correlation coefficient after exclusion of normally distributed data by Kolmogorov-Smirnov-Test and Shapiro-Wilk-Test. BL: baseline, CIS: clinically isolated syndrome FU2: two year follow-up, Gd: gadolinium enhancing lesions, OCB: oligoclonal bands, RRMS: relapsing remitting multiple sclerosis, T2: lesions in T2 weighted MRI scans, sNfL: serum neurofilament, MSFC: Multiple Sclerosis Functional Composite, YI: Youden's index, SVM: support vector machine. $^{**}p < 0.01$, $^{***}p < 0.001$.



G

	sensitivity	specificity	accuracy
OCBs + Gd⁺ lesions	71.9 %	75.9%	79.4 %
OCBs + Gd⁺ lesions + sNfL (90th percentile)	73.0 %	79.0 %	84.3 %

Fig. 2. Application of 2017 McDonald criteria selects patients with increased neuroaxonal damage. **A)** Flowchart of the study design. **B)** RRMS[2010] patients (n= 445) had significantly higher sNfL levels than CIS[2010] patients (12.5 pg/ml, IQR 8.0-22.8 pg/ml, n = 445; 10.1 pg/ml, IQR 6.9-17.1 pg/ml, n = 369, $p < 0.0005$). **C)** Application of 2017 McDonald criteria to CIS[2010] patients (n = 369) differentiated CIS[2017] (n = 45) from RRMS[2017] (n = 213) patients and resulted in significantly higher sNfL levels in patients reclassified to RRMS[2017] (10.8 pg/ml, IQR 7.4-20.1 pg/ml, n = 213) compared to patients remaining CIS[2017] (9.1 pg/ml, IQR 6.2-13.7 pg/ml, n = 45, p = 0.036 determined by Mann–Whitney-U test and 93.0% accuracy of discrimination computed by Bayesian analysis with caution of the imbalanced sample size). **D)** sNfL levels were significantly higher in RRMS[2010] patients with positive OCB (13.1 pg/ml, IQR 8.0-24.3 pg/ml, n = 222) compared to those without (10.1 pg/ml, IQR 6.7-15.8 pg/ml, n = 23, p = 0.035, discrimination accuracy = 96.4%). **E)** Significantly higher sNfL levels were found in RRMS[2010] patients with Gd+ lesions (16.4 pg/ml, IQR 9.3-31.3 pg/ml, n = 213) than in those without (10.2 pg/ml, IQR 7.2-16.1 pg/ml, n = 218, $p < 0.0005$). **F)** ROC curves generated for either positive OCB or Gd+ lesion alone (AUC = 0.76, CI 0.70 - 0.83) or in addition to the 90th percentile of sNfL (AUC = 0.84, CI 0.79 - 0.89). AUCs were compared using a Chi-square test (p = 0.035). **G)** Predictive analysis using SVM was performed yielding a prediction accuracy of 79% for OCBs

neuroaxonal injury also transiently increases NfL levels. A significantly higher proportion of patients without therapy (40%, 71/176) had higher sNfL values after two years compared to baseline in contrast to only 31% (120/392, $p < 0.05$) on "basic", 24% (32/134, $p < 0.05$) on "moderate" and 16% (17/107, $p < 0.05$) on "high" therapy (Fig. 3F). This also indicates that longitudinal sNfL changes rather than absolute sNfL values at a given time point might be indicative of disease activity and treatment stratification. Most interestingly, we found a remarkable impact of diagnostic classification on real-world treatment stratification. Compared to CIS [2010] patients, fewer RRMS[2010] patients were on "basic" treatment (RRMS[2010]: 53.7%, CIS[2010]: 72.4%), but received "moderate" (RRMS [2010]: 25.4%, CIS[2010]: 15.8%) or highly active treatment (RRMS [2010]: 20.9%, CIS[2010]: 11.8%) more frequently at year two (Fig. 3G). Importantly, evaluation of CIS[2010] patients showed that all patients but one on high therapy would have been classified as RRMS[2017] according to the new criteria (Fig. 3H). These data demonstrate that patients with initially high sNfL levels ended up on more efficient therapeutic agents two years later.

Where available, we analyzed current treatment decision at the four-year follow-up and the number of therapy changes made up to that point (no DMT: $n = 137$; "basic": $n = 209$; "moderate": $n = 132$; "high": $n = 120$, Fig. 4A). Intriguingly, without knowledge of sNfL levels, high efficacy DMT ("high" or "moderate") was initiated more often in patients with higher sNfL levels both at baseline (13.2 pg/ml, IQR 8.0–24.2 pg/ml, $n = 304$; 9.8 pg/ml, IQR 6.7–15.4 pg/ml, $n = 301$; $p < 0.0005$, Fig. 4B) and two-year follow-up (8.2 pg/ml, IQR 6.0–11.8 pg/ml, $n = 304$; 7.6 pg/ml, IQR 5.9–11.0 pg/ml, $n = 301$; $p = 0.007$). Being treated with at least one higher efficacy DMT was associated with a stronger relative decrease in sNfL levels compared to patients never treated with more than "basic" DMT ($p < 0.0005$). Furthermore, current sNfL levels at baseline or two-year follow up were significantly correlated to the number of therapy changes implemented in the subsequent two years of the observation period (sNfL at baseline $r = 0.179$, $p < 0.0005$, Fig. 4C upper panel; sNfL at two-year follow-up $r = 0.133$, $p = 0.001$, Fig. 4C lower panel).

Moreover, sNfL levels at two-year follow-up are increased in patients undergoing escalation therapy within the next two years compared to patients who stayed on the same therapy ("escalation": 9.1 pg/ml, IQR 6.4-13.7 pg/ml, $n = 107$; "no escalation": 7.7 pg/ml, IQR 5.9–11.2 pg/ml, $n = 358$, $p = 0.001$, Fig. 4D). On closer analysis, the sNfL levels changed significantly differently over time, depending on whether DMT was initiated/escalated, de-escalated or maintained in the same class between two and four years of follow-up ($p = 0.029$). The decision for "no escalation" was retrospectively confirmed by lower sNfL levels at two-year follow-up compared to patients with subsequent DMT initiation/escalation (7.1 pg/ml, IQR 5.8–10.4 pg/ml, $n = 50$; 9.1 pg/ml, IQR 6.4–13.7 pg/ml, $n = 107$, $p = 0.026$, Fig. 4D). Since different treatment strategies, namely early induction or later escalation to higher efficacy DMT, are part of the current debate we next separated all patients ending up in DMT groups "moderate" or "high" at four-year follow-up on whether they were already in these DMT classes at two-year follow up ("induction": $n = 153$) or not ("escalation", $n = 82$, Fig. 4E). Supporting the potential of neurofilament as a treatment biomarker we found a significant interaction between treatment strategy and time on sNfL levels ($p = 0.002$). Although patients in the "induction" group initially started with higher baseline sNfL levels (14.4 pg/ml, IQR 8.2–28.9 pg/ml; 10.9 pg/ml, IQR 7.5–19.1 pg/ml, $p = 0.035$), it declined much more sharply in the first two years and even crossed the line of the "escalation" group (7.8 pg/ml, IQR 6.0–12.3 pg/ml; 9.3 pg/ml, IQR 6.3–14.3 pg/ml, $p = 0.025$). Nonetheless, the escalated therapeutic regimen between year two and four was reflected by a stronger sNfL level decrease leading to similar four-year follow-up sNfL levels in both groups ("induction": 6.2 pg/ml, IQR 4.8–8.7 pg/ml; "escalation": 5.7 pg/ml, IQR 4.4–7.9 pg/ml, $p = 0.885$). These data underline that longitudinal assessment of sNfL levels might be a suitable approach for a real-world comparison of different treatment stratification algorithms.

4. Discussion

We here evaluated sNfL in a large cohort of early MS patients and provide supporting evidence for a role of starting longitudinal sNfL assessments directly at the time point of first diagnosis. Specifically, sNfL measurements have implications for i) diagnostic accuracy, ii) prognosis, and iii) treatment decision making in the four years after diagnosis.

In their 2017 revision of the McDonald diagnostic criteria, an international panel of experts reached a consensus that makes it easier to establish a diagnosis of MS earlier than was possible with previous criteria. All CIS[2010] patients but one under high efficacy therapies ($n = 22$) at follow-up would have already been diagnosed as "RRMS" according to the 2017 criteria, providing robust data for the superior diagnostic value of the 2017 diagnostic criteria. Previously, it was reported that increased CSF NfL in patients with radiologically isolated syndrome is an independent risk factor of developing CIS [29] and for further development of clinically definite MS in CIS patients [30]. Our findings even show that sNfL might be useful in differentiating CIS from RRMS and may thus be considered as a parameter for future revisions of diagnostic criteria if sNfL methology can be robustly improved to allow for clinical routine care settings. In fact, highest sNfL levels (in our patient cohort: cutoff 90th percentile) increased sensitivity, specificity, and accuracy over OCB or symptomatic Gd+ lesions to discriminate between CIS and MS. Further studies should specifically focus on an additional diagnostic value in patients with first disease symptoms and high initial sNfL levels and thus a high likelihood of axonal damage and an established CNS autoimmune inflammation. This notion is further supported by our data showing that the risk predicting EDSS/MSFC after two years based on baseline T2 lesions and Gd+ lesions was markedly elevated by additionally including sNfL values.

In addition to a potential added value in the initial diagnosis of patients, sNFL might also serve as a marker of treatment response, as sNFL levels have been described to decrease after initiation of any DMT [8,10,20] and specifically after switch from injectable therapies to fingolimod [20] or after initiation of interferon-beta-treatment [31]. We here present findings from one of the so far largest and earliest cohorts of MS patients correlating treatment responses with sNfL levels. Importantly, data was acquired in a prospective, centralised and highly standardised manner. The number of patients without DMT after two years is comparable with other databases (e.g., MSBase ~30% of patients; [32] Swiss National Multiple Sclerosis Cohort ~50% of patients [10]). However, patients in our study were recruited at first demyelinating event and therefore earlier than in other studies. After a two-year follow-up, untreated patients had an elevated risk of higher sNfL levels than patients on DMT. It should be noted that mean sNfL levels were nevertheless reduced in all patient groups, and to a lesser extent in untreated patients, after two and four years compared to baseline. Previously, sNfL levels were found to be elevated 2 months before and 1 month after MRI scans showing Gd+ lesions [31]. Inclusion criteria for our cohort demanded a minimum interval of 30 days after relapse, but the exact duration of elevated sNfL due to clinical or MRI disease activity has not yet been established. This underlines the importance of both longitudinal

and/or Gd+ for discriminating CIS and RRMS according to 2017 McDonald criteria. Accuracy increased to 84% by including the 90th percentile of sNfL in addition to the above two variables. Group differences were analyzed by Mann–Whitney-U test and in cases of imbalanced sample sizes additionally validated by Bayesian analysis (C+D). sNfL levels are reported as median. IQR: interquartile range, CIS: clinically isolated syndrome, DIS: dissemination in space, DIT: dissemination in time, Gd: gadolinium, OCB: oligoclonal bands, sNfL: serum neurofilament, RRMS: relapsing-remitting multiple sclerosis, SVM: support vector machine algorithm, ROC: receiver operating characteristic, AUC: area under the curve. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns = not significant.

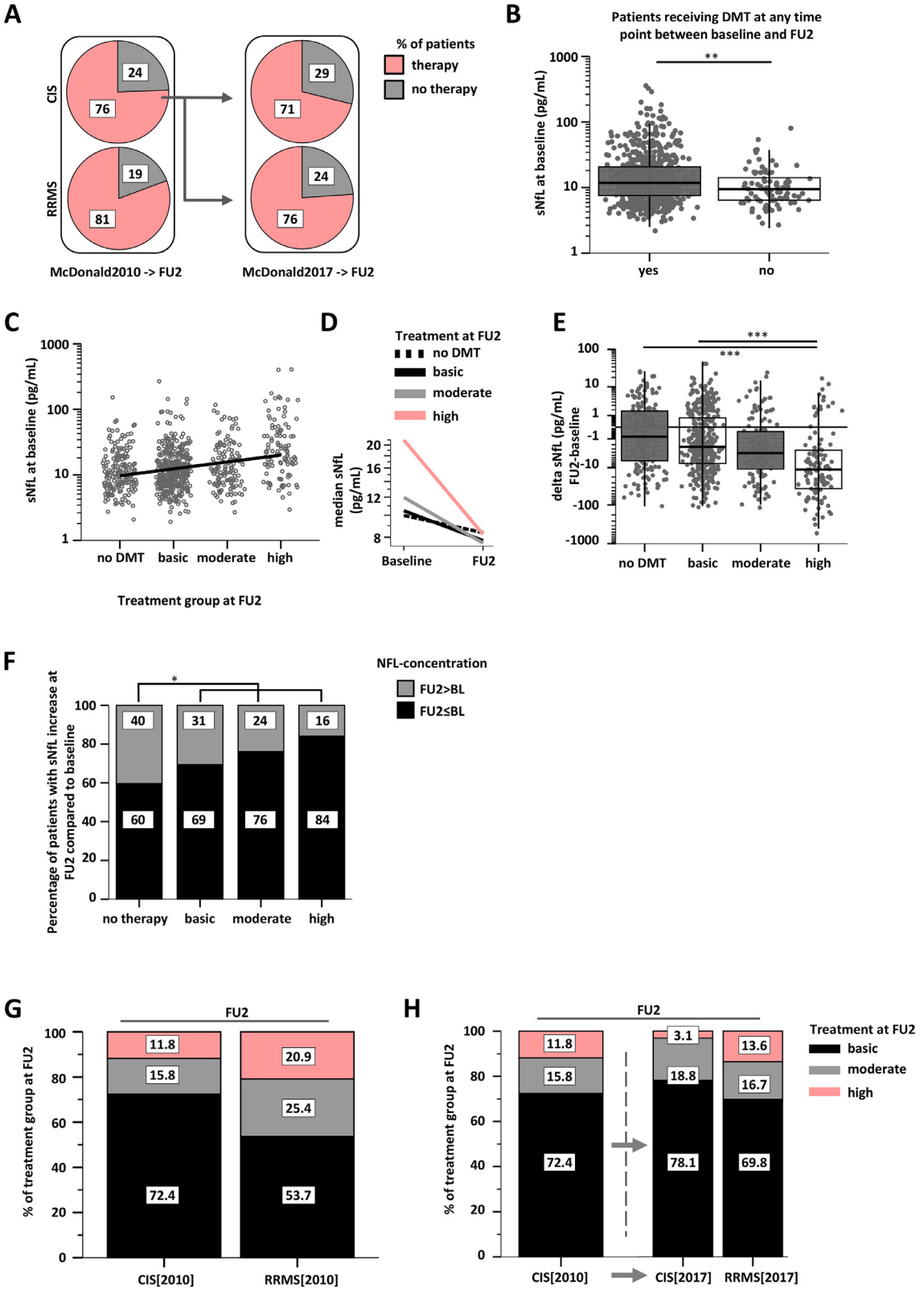


Fig. 3. Relationship between 2017 McDonald criteria, therapeutic decision and sNfL levels at two-year follow-up. **A)** Percentage of patients receiving DMT in CIS[2010], RRMS[2010], CIS[2010]→CIS[2017] and CIS[2010]→RRMS[2017] at two-year follow-up. **B)** sNfL levels at baseline were significantly higher in patients receiving DMT in the following two years (11.8 pg/ml, IQR 7.5-20.9 pg/ml, $n = 727$) than patients without specific MS medication until two-year follow-up (9.5 pg/ml, IQR 6.4-14.1 pg/ml, $n = 87$, $p = 0.002$). **C)** Patients were grouped into four classes based on the type of DMT they were receiving at year two follow-up. Median sNfL levels at baseline and two-year follow-up in four patient groups defined above (no DMT: $n = 176$, basic: $n = 392$, moderate: $n = 134$ and high: $n = 134$). Baseline sNfL levels were correlated with the patient's treatment group at two-year follow-up ($r = 0.223$, $p < 0.0005$). **D)** There was a statistically significant two-way interaction regarding the between- and within-subject factors (time*treatment groups) on sNfL concentration ($p < 0.0005$). At baseline the high treatment group showed elevated sNfL concentration compared to all other groups ($p < 0.0005$). We observed no significant difference between the treatment groups at two-year follow-up (ns) but a statistically significant effect of time on sNfL concentration for all four groups ($p < 0.0005$). **E)** Delta sNfL (two-year follow-up minus baseline) for patients without DMT and on basic, moderate and high DMT. Calculation was adjusted for baseline sNfL values. High treatment versus basic treatment

measurements and interpretation of sNfL results together with corresponding clinical information. Strikingly, patients who were on high efficacy therapies after two years, reflecting a severe early disease course, had the highest initial sNfL levels and strongest relative decrease. Furthermore, treatment decisions and escalation/de-escalation decisions within the first four years were reflected by changes in sNfL values. Future studies should further evaluate whether sNfL levels are suited for prospective treatment stratification. Furthermore, to pave the way into clinical practice, age-, demographic-, and comorbidity-associated normative values as well as optimal frequency of sampling and thresholds leading to distinct clinical decisions in patient management, need to be internationally agreed on and confirmed applying the exact same protocol prior to inter-laboratory comparisons and accreditation. It should, however, be mentioned that at present, there is in parallel still work to do with regards to missing mechanistic understanding about the underlying pathophysiologic processes best reflected by sNfL. How do inflammatory versus "diffuse" neurodegenerative processes impact sNfL both on a short- and long-term scale, and what would be the add-on value to use different surrogate parameters all considered to reflect neurodegenerative processes in MS (e.g., brain atrophy, OCT, sNfL).

Taken together, we report here findings in a large German multicentre cohort with early CIS/RRMS. In a clinical setting, determining sNfL levels at time point of diagnosis and thereafter longitudinally might not only increase the sensitivity of diagnostic criteria, but could also – at least according to our findings based on the NationMS cohort and German treatment practice in expert centres – provide the next step towards personalised and optimised MS therapy.

Funding sources

The German NationMS cohort and KKNMS were supported by grants from the German Federal Ministry for Education and Research (BMBF) Grant nos. 01GI0914, 01GI1601I (Bochum); 01GI0916, 01GI1601G (Lübeck); and 01GI1601B (Marburg). The study was supported by the Biobank of the Department of Neurology as part of the Joint Biobank Munich in the framework of the German Biobank Node, the German Research Council (DFG, CRC-TR-128 to S.G., F.Z., and S.B.), and Hertie-Stiftung (to F.S. and S.B.). The funding sources played no role in the writing of the manuscript or the decision to submit it for publication. The authors have not been paid to write this article by a pharmaceutical company or other agency. The corresponding author had full access to all data and had final responsibility for the decision to submit the manuscript for publication.

Declaration of Competing Interest

Falk Steffen, Vinzenz Fleischer, Muthuraman Muthuraman, Sergiu Groppa, and Mark Mühlau declare no competing interests. Stefan Bittner has received honoria and compensation for travel from Biogen Idec, Merck Serono, Novartis, Sanofi-Genzyme and Roche. Timo Uphaus received honoria from Merck Serono. Carsten Lukas received consulting and speaker's honoraria from BiogenIdec, Bayer Schering, Novartis, Sanofi, Genzyme, and TEVA; has received research scientific grant support from Bayer Schering, TEVA, and Merck Serono; holds an endowed professorship supported by the Novartis Foundation. Anke Salmen received speaker honoraria and/or travel compensation for activities with Almirall Hermal GmbH, Biogen, Merck, Novartis, Roche, and Sanofi Genzyme and research support by the Swiss MS Society, none related to this work. Felix Luessi received consultancy fees from Roche and support with travel cost from Teva Pharma. Achim Berthele reports personal fees from Bayer Healthcare, Biogen, Merck Serono, Mylan, Roche, and Sanofi Genzyme, and his institution received compensations for clinical trials from Alexion Pharmaceuticals, Biogen, Chugai, Novartis, Roche, Sanofi Genzyme, and Teva - all outside the submitted work. Luisa Klotz received honoraria for lecturing and serving on advisory boards, as well as travel expenses for attending meetings and financial research support from Immunic AG, Biogen, Janssen, Merck Serono, Novartis, Roche, Sanofi Genzyme, Teva, Grifols, Alexion, Santhera, Bayer Healthcare, the Deutsche Forschungsgemeinschaft (DFG; German Research Society), the German Ministry for Education and Research (BMBF), the Interdisciplinary Center for Clinical Studies (IZKF) Muenster and the program Innovative Medical Research (IMF) Muenster. Sven G. Meuth receives honoraria for lecturing, and travel expenses for attending meetings from Almirall, Amicus Therapeutics Germany, Bayer Health Care, Biogen, Celgene, Diamed, Genzyme, MedDay Pharmaceuticals, Merck Serono, Novartis, Novo Nordisk, ONO Pharma, Roche, Sanofi-Aventis, Chugai Pharma, QuintilesIMS and Teva. His research is funded by the German Ministry for Education and Research (BMBF), Deutsche Forschungsgemeinschaft (DFG), Else Kröner Fresenius Foundation, German Academic Exchange Service, Hertie Foundation, Interdisciplinary Center for Clinical Studies (IZKF) Muenster, German Foundation Neurology and Almirall, Amicus Therapeutics Germany, Biogen, Diamed, Fresenius Medical Care, Genzyme, Merck Serono, Novartis, ONO Pharma, Roche, and Teva. Antonios Bayas received personal compensation from Merck, Biogen, Bayer Vital, Novartis, TEVA, Roche and Sanofi-Aventis/Genzyme and grants for congress trips and participation from Biogen, TEVA, Novartis, Sanofi-Aventis/Genzyme, and Merck. Friedemann Paul receives honoraria for lecturing, and travel expenses for attending meetings from Guthy Jackson Foundation, Sanofi Genzyme, Novartis, Alexion, Viela Bio, Roche, UCB, Mitsubishi Tanabe and Celgene. His research is funded by the German Ministry for Education and Research (BMBF), Deutsche Forschungsgemeinschaft (DFG), Einstein Foundation, Guthy Jackson Charitable Foundation, EU FP7 Framework Program, Arthur Arnstein Foundation Berlin, Biogen, Genzyme, Merck Serono, Novartis, Bayer, Teva, Alexion, Roche, Parexel and Almirall. Hans-Peter Hartung has received fees for consulting, serving on steering committees and data monitoring committees from Bayer Healthcare, Biogen, GeNeuro, Medimmune, Merck, Novartis, Roche, Sanofi Genzyme, Teva and TG Therapeutics with approval by the Rector of Heinrich-Heine-University. Ralf Linker received Research Support and/or personal compensation for activities with Bayer Health Care, Biogen, Genzyme/Sanofi, Merck, Novartis Pharma, Roche, and TEVA Pharma; none related to this work. Christoph Heesen received research grants and speaker honoraria from Biogen, Genzyme, Roche, and Merck; none related to this work. Martin Stangel received honoraria for scientific lectures or consultancy from Alexion, Bayer Healthcare, Biogen, CSL Behring, Grifols, Janssen, Merck-Serono, Novartis, Roche, Sanofi-Aventis, Takeda, and Teva. His

($p = 0.001$) and no DMT ($p = 0.001$) were significantly different. **F**) sNfL levels were compared between baseline (BL) and two-year follow-up (FU2). Percentage of patients with sNfL increase (FU2>BL) and sNfL decrease (FU2<BL) are depicted for four treatment groups. The group without treatment showed significantly higher proportion of patients with increased sNfL levels as compared to the groups with either moderate or high treatment. **G)** Percentage of patients under basic, moderate and high efficacy DMT in CIS[2010] and RRMS[2010] after two-year follow-up. **H)** Percentage of patients under basic, moderate and high efficacy DMT in CIS[2010] patients and in CIS[2017] and RRMS[2017] patients after reassessment according to 2017 criteria. Group differences were analyzed by Mann–Whitney-U test (B) or a one-way ANCOVA with baseline sNfL values as a covariate (E). For two-way (treatment group*time) interaction a two-way mixed ANOVA with four additional separate one-way ANOVAs with Tukey post-hoc test for multiple comparison for simple main effects/between-subject factors for the different treatment groups was applied (D). Correlation analysis was performed by Spearman's rank correlation coefficient after exclusion of normally distributed data by Kolmogorov–Smirnov-Test and Shapiro–Wilk-Test. Chi-square test of homogeneity was applied to investigate differences in proportions. Post hoc analysis involved pairwise comparison using multiple z-test of two proportions and, where appropriate, a Bonferroni correction. sNfL levels are reported as median. FU2: two-year follow-up, CIS: clinically isolated syndrome, RRMS: relapsing remitting multiple sclerosis, DMT: disease-modifying therapy, basic: interferons and glatirameracetate; moderate: teriflunomide and dimethylfumarate; high: natalizumab, rituximab, fingolimod, ocrelizumab, daclizumab, alemtuzumab, mitoxantrone. ns: not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

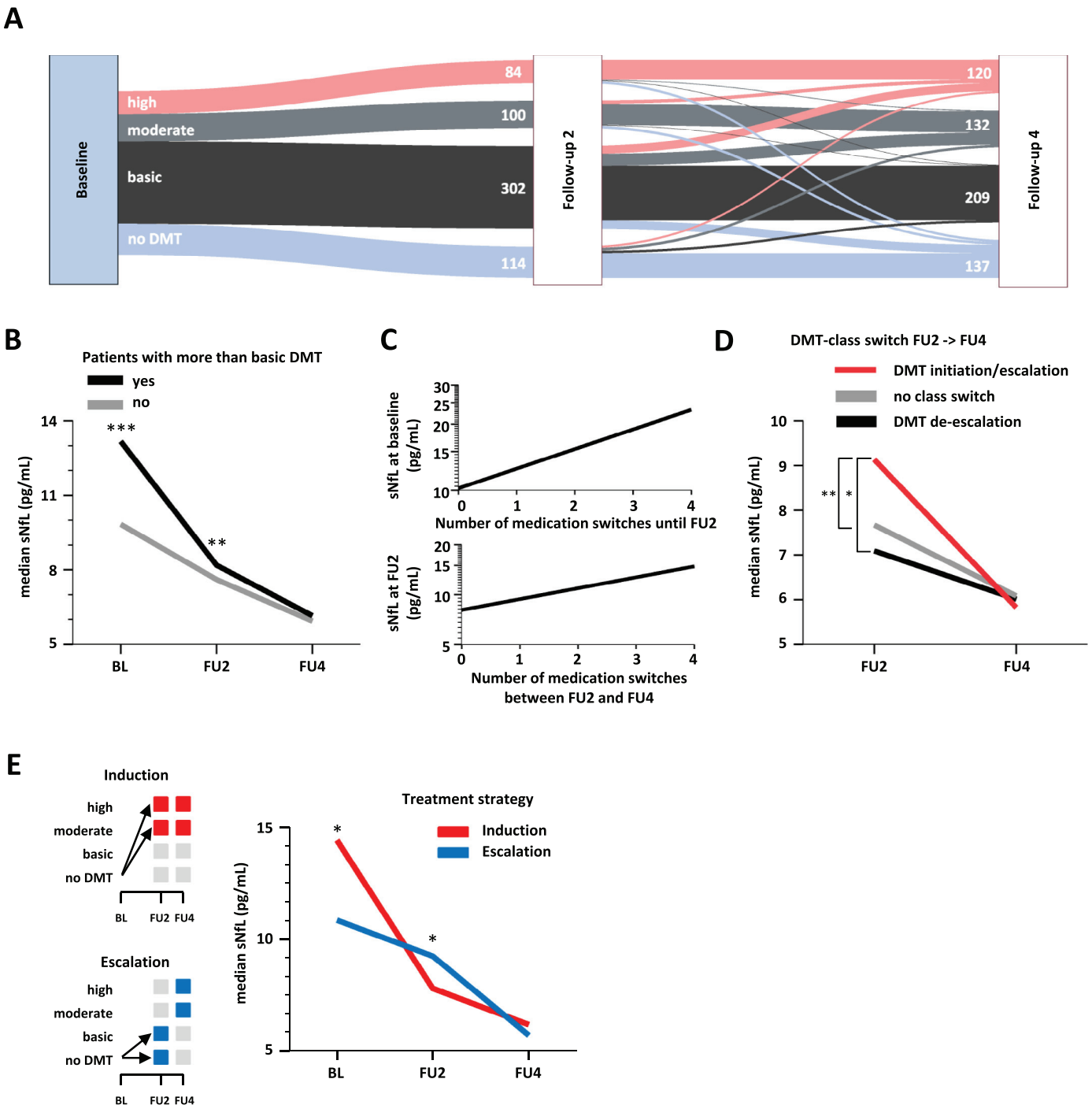


Fig. 4. sNfL levels reflect therapeutic decisions within four years after diagnosis. **A)** Sankey diagram illustrating treatment stratification at two-year follow-up and therapy changes between year two and four. Width of arrows reflects quantitative changes between groups. **B)** High efficacy DMTs ("high" or "moderate") were initiated more often in patients with higher sNfL levels both at baseline (13.2 pg/ml, IQR 8.0-24.2 pg/ml, n = 304; 9.8 pg/ml, IQR 6.7-15.4 pg/ml, n = 301; $p < 0.0005$) and two-year follow-up (8.2 pg/ml, IQR 6.0-11.8 pg/ml, n = 304; 7.6 pg/ml, IQR 5.9-11.0 pg/ml, n = 301; $p = 0.007$). There was a statistically significant two-way interaction on sNfL concentration ($p < 0.0005$). **C)** Top: baseline sNfL levels were correlated with the number of medication switches performed between baseline and two-year follow-up ($r = 0.179$, $p < 0.0005$). Below: FU2 sNfL levels were correlated with the number of medication switches performed between two-year and four-year follow-up ($r = 0.133$, p = 0.001). **D)** sNfL at two-year follow-up was significantly increased in patients with escalation of DMT treatment group between two-year and four-year follow-up compared to patients without escalation of DMT within this period (9.1 pg/ml, IQR 6.4-13.7 pg/ml, n = 107; 7.7 pg/ml, IQR 5.9-11.2 pg/ml, n = 358, $p = 0.001$). **E)** Patients with moderate or high DMT group at four-year follow-up were divided in two different groups depending on whether they have already been in these DMT classes at two-year follow up ("induction": n = 153) or not ("escalation": n = 82). A schematic illustration of this process is depicted in the left panel. Compared to the "escalation" group patients in the "induction" group showed higher sNfL levels at baseline (14.4 pg/ml, IQR 8.2-28.9 pg/ml; 10.9 pg/ml, IQR 7.5-19.1 pg/ml, $p = 0.035$) but lower sNfL concentrations at two-year follow-up (7.8 pg/ml, IQR 6.0-12.3 pg/ml; 9.3 pg/ml, IQR 6.3-14.3 pg/ml, $p = 0.025$). There was no significant difference at four-year follow-up ("induction": 6.2 pg/ml, IQR 4.8-8.7 pg/ml; "escalation": 5.7 pg/ml, IQR 4.4-7.9 pg/ml, $p = 0.885$). Correlation analysis was performed by Spearman's rank correlation coefficient after exclusion of normally distributed data by Kolmogorov-Smirnov-Test and Shapiro–Wilk-Test. Where appropriate, post hoc analysis was performed using a Bonferroni correction. sNfL levels are reported as median and interquartile range (IQR). FU2: two-year follow-up, FU4: four-year follow-up, DMT: disease-modifying therapy, basic: interferons and glatirameracetate; moderate: teriflunomide and dimethylfumarate; high: natalizumab, rituximab, fingolimod, ocrelizumab, daclizumab, alemtuzumab, mitoxantrone. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

institution received research support from Bayer Healthcare, Biogen Idec, Genzyme, Merck-Serono, Novartis, and Teva; none related to this work. Brigitte Wildemann received grants from the German Ministry of Education and Research, Deutsche Forschungsgemeinschaft, Dietmar Hopp Foundation and Klaus Tschira Foundation, grants and personal fees from Merck Serono, Sanofi Genzyme, Novartis pharmaceuticals, and personal fees from Bayer Healthcare, Biogenm, Teva Pharma; none related to this work. Florian Then Bergh received travel support to attend scientific meetings, personal speaker honoraria, and consultancy fees as a speaker and advisor from Bayer Healthcare, Biogen, Merck Serono, Novartis, Roche, Sanofi Genzyme, and TEVA. He received, through his institution, unrestricted research grants from Novartis, TEVA, Bayer Healthcare, and Actelion; none related to this work. He received funding from the DFG and, through TRM Leipzig, from the BMBF. Björn Tackenberg received personal speaker honoraria and consultancy fees as a speaker and advisor from Bayer Healthcare, Biogen, CSL Behring, GRIFOLS, Merck Serono, Novartis, Octapharma, Roche, Sanofi Genzyme, TEVA und UCB Pharma. His University received unrestricted research grants from Biogen-idec, Novartis, TEVA, Bayer Healthcare, CSL-Behring, GRIFOLS, Octapharma, Sanofi Genzyme und UCB Pharma; none related to this work. He is currently an employee of Roche. The data collection, evaluation and drafting of the manuscript was performed before beginning employment at Roche. Tania Kuempfel has received travel expenses and personal compensations from Bayer Healthcare, Teva Pharma, Merck-Serono, Novartis, Sanofi-Aventis/Genzyme, Roche and Biogen as well as grant support from Bayer-Schering AG, Novartis and Chugai Pharma. Frank Weber received honoraria from Genzyme, Novartis TEVA and Biogen for speaking or for serving on a scientific advisory board, a travel grant for the attention of a scientific meeting from Merck-Serono and Novartis and grant support from Merck-Serono, Novartis and from the Federal Ministry of Education and Research (BMBF, Projects Biobanking and Omics in ControlMS as part of the Competence Network Multiple Sclerosis). Uwe K. Zettl received speaker fees, travel compensation and/or his section received research support from Alexion, Almirall, Bayer Health Care, Biogen, Celgene, Genzyme, Merck Serono, Novartis, Roche, Sanofi-Aventis, Teva and grants from German Ministry for Education and Research (BMBF), German Ministry for Economy (BMWi), Deutsche Forschungsgemeinschaft (DFG), European Union (EU), outside the submitted work. Ulf Ziemann received speaker honoraria and/or travel compensation from Biogen Idec GmbH, Bayer Vital GmbH, Bristol Myers Squibb GmbH, CorTec GmbH, Medtronic GmbH, and grants from Biogen Idec GmbH, Servier, and Janssen Pharmaceuticals NV; none related to this work. Hayrettin Tumani received speaker honoraria from Bayer, Biogen, Fresenius, Genzyme, Merck, Novartis, Roche, Siemens, Teva; serves as section editor for the Journal of Neurology, Psychiatry, and Brain Research; and his institution receives research support from Fresenius, Genzyme, Merck, and Novartis; none related to this work. Bernhard Hemmer served on scientific advisory boards for Novartis; he has served as DMSC member for AllergyCare, Polpharma, and TG therapeutics; he or his institution have received speaker honoraria from Desitin; his institution has received research support from Regeneron; holds part of two patents; one for the detection of antibodies and T cells against KIR4.1 in a subpopulation of MS patients and one for genetic determinants of neutralising antibodies to interferon β during the last 3 years. Heinz Wiendl receives honoraria for acting as a member of Scientific Advisory Boards and as a consultant for Biogen, Evgen, MedDay Pharmaceuticals, Merck Serono, Novartis, Roche Pharma AG, Sanofi-Genzyme, as well as speaker honoraria and travel support from Alexion, Biogen, Cognomed, F. Hoffmann-La Roche Ltd., Gemeinnützige Hertie-Stiftung, Merck Serono, Novartis, Roche Pharma AG, Sanofi-Genzyme, TEVA, and WebMD Global. Prof. Wiendl is acting as a paid consultant for Abbvie, Actelion, Biogen, IGES, Novartis, Roche, Sanofi-Genzyme, and the Swiss Multiple Sclerosis Society. His research is funded by the BMBF, DFG, Else Kröner Fresenius Foundation, Fresenius Foundation, Hertie Foundation, NRW Ministry of Education and Research, Interdisciplinary Center for Clinical Studies (IZKF) Muenster and RE Children's Foundation, Biogen GmbH, GlaxoSmithKline GmbH, and Roche Pharma AG, Sanofi-Genzyme. Ralf Gold serves on scientific advisory boards for Teva Pharmaceutical Industries Ltd., Biogen Idec, Bayer Schering Pharma, and Novartis; has received speaker honoraria from Biogen Idec, Teva Pharmaceutical Industries Ltd., Bayer Schering Pharma, and Novartis; serves as editor for Therapeutic Advances in Neurological Diseases and on the editorial boards of Experimental Neurology and the Journal of Neuroimmunology; and receives research support from Teva Pharmaceutical Industries Ltd., Biogen Idec, Bayer Schering Pharma, Genzyme, Merck Serono, and Novartis; none related to this work. Frauke Zipp has recently received research grants and/or consultation funds from the DFG, BMBF, PMSA, Novartis, Octapharma, Merck Serono, ONO Pharma, Biogen, Genzyme, Celgene and Roche.

Acknowledgements

The authors and representatives of the KKNMS express their gratitude to all contributors of the study, especially the study nurses, for their motivated collaboration and recruitment efforts, all the patients and relatives for their participation and support, and the data monitoring and administrative personnel of the study. We thank Gisela Antony (Central Information Office, Marburg) for excellent IT support of the KKNMS cohort. The authors thank Rosie Gilchrist and Cheryl Ernest for proofreading and editing the manuscript.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2020.102807.

References

[1] Larochelle C, Uphaus T, Prat A, Zipp F. Secondary progression in multiple sclerosis: neuronal exhaustion or distinct pathology? Trends Neurosci 2016;39(5):325–39.
[2] Saidha S, Al-Louzi O, Ratchford JN, et al. Optical coherence tomography reflects brain atrophy in multiple sclerosis: a four-year study. Ann Neurol 2015;78(5):801–13.
[3] Zimmermann HG, Knier B, Oberwahrenbrock T, et al. Association of retinal ganglion cell layer thickness with future disease activity in patients with clinically isolated syndrome. JAMA Neurol 2018;75(9):1071–9.
[4] Steenwijk MD, Geurts JJ, Daams M, et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain: J Neurol 2016;139(Pt 1):115–26.
[5] Khalil M, Teunissen CE, Otto M, et al. Neurofilaments as biomarkers in neurological disorders. Nat Rev Neurol 2018;14(10):577–89.
[6] Preische O, Schultz SA, Apel A, et al. Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease. Nat Med 2019;25(2):277–83.
[7] Gattringer T, Pinter D, Enzinger C, et al. Serum neurofilament light is sensitive to active cerebral small vessel disease. Neurology 2017;89(20):2108–14.
[8] Siller N, Kuhle J, Muthuraman M, et al. Serum neurofilament light chain is a biomarker of acute and chronic neuronal damage in early multiple sclerosis. Mult Scler 2018;25(5):678–86.
[9] Barro C, Benkert P, Disanto G, et al. Serum neurofilament as a predictor of disease worsening and brain and spinal cord atrophy in multiple sclerosis. Brain: J Neurol 2018;141(8):2382–91.
[10] Disanto G, Barro C, Benkert P, et al. Serum Neurofilament light: a biomarker of neuronal damage in multiple sclerosis. Ann Neurol 2017;81(6):857–70.
[11] Canto E, Barro C, Zhao C, et al. Association between serum neurofilament light chain levels and long-term disease course among patients with multiple sclerosis followed up for 12 years. JAMA Neurol 2019 [Epub ahead of print]. doi: 10.1001/jamaneurol.2019.2137.
[12] Wong YYM, Bruijstens AL, Barro C, et al. Serum neurofilament light chain in pediatric MS and other acquired demyelinating syndromes. Neurology 2019;93(10):e968–e74.
[13] Bjornevik K, Munger KL, Cortese M, et al. Serum Neurofilament Light Chain Levels in Patients With Presymptomatic Multiple Sclerosis. JAMA Neurol 2019 [Epub ahead of print]. doi: 10.1001/jamaneurol.2019.3238.
[14] Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018;17(2):162–73.

[15] Filippi M, Rocca MA, Ciccarelli O, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol 2016;15(3):292–303.
[16] Beesley R, Anderson V, Harding KE, et al. Impact of the 2017 revisions to McDonald criteria on the diagnosis of multiple sclerosis. Mult Scler 2018;24(13):1786–7.
[17] McNicholas N, Lockhart A, Yap SM, et al. New versus old: Implications of evolving diagnostic criteria for relapsing-remitting multiple sclerosis. Mult Scler 2019;25(6):867–70.
[18] Schwenkenbecher P, Wurster U, Suhs KW, Stangel M, Skripuletz T. Applying the 2017 McDonald diagnostic criteria for multiple sclerosis. Lancet Neurol 2018;17(6):498.
[19] von Bismarck O, Dankowski T, Ambrosius B, et al. Treatment choices and neuropsychological symptoms of a large cohort of early MS. Neurol(R) Neuroimmunol Neuroinflamm 2018;5(3):e446.
[20] Piehl F, Kockum I, Khademi M, et al. Plasma neurofilament light chain levels in patients with MS switching from injectable therapies to fingolimod. Mult Scler 2018;24(8):1046–54.
[21] Cutter GR, Baier ML, Rudick RA, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain: J Neurol 1999;122(Pt 5):871–82.
[22] Johnen A, Burkner PC, Landmeyer NC, et al. Can we predict cognitive decline after initial diagnosis of multiple sclerosis? Results from the German National early MS cohort (KKNMS). J Neurol 2019;266(2):386–97.
[23] Fischer JS, Rudick RA, Cutter GR, Reingold SC. The multiple sclerosis functional composite measure (MSFC): an integrated approach to MS clinical outcome assessment. national MS society clinical outcomes assessment task force. Mult Scler 1999;5(4):244–50.
[24] Khalil M, Pirpamer L, Hofer E, et al. Normative values of serum neurofilament light levels. Mult Scler 2018;24(S(2)):530–737.
[25] Kruschke JK. Bayesian estimation supersedes the t test. J Exp Psychol Gen 2013;142(2):573–603.
[26] Schuberth F, Henseler J, Dijkstra TK. Partial least squares path modeling using ordinal categorical indicators. Qual Quant 2018;52(1):9–35.
[27] Barkhof F, Filippi M, Miller DH, et al. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. Brain: J Neurol 1997;120(Pt 11):2059–69.
[28] Thompson AJ, Reingold SC, Cohen JA. International panel on diagnosis of multiple S. Applying the 2017 McDonald diagnostic criteria for multiple sclerosis – authors' reply. Lancet Neurol 2018;17(6):499–500.
[29] Matute-Blanch C, Villar LM, Alvarez-Cermeno JC, et al. Neurofilament light chain and oligoclonal bands are prognostic biomarkers in radiologically isolated syndrome. Brain: J Neurol 2018;141(4):1085–93.
[30] Teunissen CE, Iacobaeus E, Khademi M, et al. Combination of CSF N-acetylaspartate and neurofilaments in multiple sclerosis. Neurology 2009;72(15):1322–9.
[31] Varhaug KN, Barro C, Bjornevik K, et al. Neurofilament light chain predicts disease activity in relapsing-remitting MS. Neurol(R) Neuroimmunol Neuroinflamm 2018;5(1):e422.
[32] Brown JWL, Coles A, Horakova D, et al. Association of Initial disease-modifying therapy with later conversion to secondary progressive multiple sclerosis. JAMA 2019;321(2):175–87.