Enhancing clinical value of serum neurofilament light chain measurement

Peter Kosa, …, Mary Sandford, Bibiana Bielekova


Serum neurofilament light chain (sNFL) is becoming an important biomarker of neuroaxonal injury. Though sNFL correlates with cerebrospinal fluid (CSF) NFL (cNFL), 40-60% of variance remains unexplained. We aimed to mathematically adjust sNFL to strengthen its clinical value. We measured NFL in blinded fashion in 1,138 matched CSF and serum samples from 571 subjects. Multiple linear regression (MLR) models constructed in the training cohort were validated in an independent cohort. MLR model that included age, blood urea nitrogen (BUN), alkaline phosphatase (AP), creatinine, and weight improved correlations of cNFL with sNFL (from R² = 0.57 to 0.67). Covariate-adjustment significantly improved the correlation of sNFL with number of contrast-enhancing lesions (from R² = 0.18 to 0.28; 36% improvement) in the validation cohort. Unexpectedly, only sNFL, but not cNFL, weakly but significantly correlated with cross-sectional MS severity outcomes. Investigating two non-overlapping hypotheses, we show that subjects with proportionally higher sNFL to cNFL have higher clinical and radiological evidence of spinal cord (SC) injury, and likely release NFL from peripheral axons into blood, bypassing the CSF. Thus, sNFL captures two sources of axonal injury: central and peripheral; the latter reflecting SC damage, which […]

Find the latest version:

https://jci.me/161415/pdf
Enhancing clinical value of serum neurofilament light chain measurement

Peter Kosa¹*, Ruturaj Masvekar¹*, Mika Komori¹$, Jonathan Phillips¹, Vignesh Ramesh¹, Mihael Varosanec¹, Mary Sandford¹ and Bibiana Bielekova¹#

¹Neuroimmunological Diseases Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

*Share first authorship, listed in alphabetical order of last names.

$ Contributed to this work as a former employee of the National Institutes of Health, the opinions expressed in this manuscript do not represent her current affiliation (Eli Lilly, Kobe, Japan).

#Corresponding author:

Bibi Bielekova, MD

10 Center Drive MSC 1400

Bldg 10, Rm 5N-248

Bethesda, MD 20892

Phone: (301) 402-4488

E-mail: Bibi.Bielekova@nih.gov

Conflict-of-interest statement: The authors have declared that no conflict of interest exists.
Abstract

Serum neurofilament light chain (sNFL) is becoming an important biomarker of neuroaxonal injury. Though sNFL correlates with cerebrospinal fluid (CSF) NFL (cNFL), 40-60% of variance remains unexplained. We aimed to mathematically adjust sNFL to strengthen its clinical value. We measured NFL in blinded fashion in 1,138 matched CSF and serum samples from 571 subjects. Multiple linear regression (MLR) models constructed in the training cohort were validated in an independent cohort. MLR model that included age, blood urea nitrogen (BUN), alkaline phosphatase (AP), creatinine, and weight improved correlations of cNFL with sNFL (from $R^2 = 0.57$ to 0.67). Covariate-adjustment significantly improved the correlation of sNFL with number of contrast-enhancing lesions (from $R^2 = 0.18$ to 0.28; 36% improvement) in the validation cohort. Unexpectedly, only sNFL, but not cNFL, weakly but significantly correlated with cross-sectional MS severity outcomes. Investigating two non-overlapping hypotheses, we show that subjects with proportionally higher sNFL to cNFL have higher clinical and radiological evidence of spinal cord (SC) injury, and likely release NFL from peripheral axons into blood, bypassing the CSF. Thus, sNFL captures two sources of axonal injury: central and peripheral; the latter reflecting SC damage, which primarily drives disability progression in MS.
**Introduction**

While there are many excellent laboratory tests reflecting injury to peripheral organs such as liver or kidneys, quantifying central nervous system (CNS) injury with peripheral blood biomarkers is extremely difficult because the concentrations of analytes released during neuronal damage are below detection levels of most of the conventional immunoassays. Two developments in recent time have changed this reality: 1) New, advanced immunoassays, such as Single Molecule Array (SIMOA®), capable of reliably measuring femtomolar concentrations of proteins; and 2) Demonstration that neurofilament light chain (NFL), a protein exclusively expressed in neurons, remains elevated for weeks after acute neuro-axonal injury. Indeed, after traumatic brain injury (TBI), the serum levels of some neuronal proteins such as ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) and neuron-specific enolase (NSE) peak at 24h and become undetectable 96 hours later, while NFL levels rise gradually, peak in several days and remain elevated for up to a few weeks (1). This unusual dynamic makes (serum/plasma) NFL an excellent biomarker of acute neuronal injuries, such as those caused by trauma or hypoxia, but also for monitoring of chronic neurological diseases, including multiple sclerosis (MS).

Many excellent papers demonstrated association between CSF and serum NFL levels and formation of acute MS lesions (i.e., contrast-enhancing lesions [CELs] on brain MRI (2, 3)) or association of therapy-induced inhibition of serum/CSF NFL concentration with therapeutic effect on CEL or MS relapses (4). These important observations raised a possibility to use serum NFL (sNFL) measurements for managing MS on a patient level.

Clinical value of the laboratory test is assessed by sensitivity (i.e., the ability of the test to correctly identify people with the measured process) and specificity (i.e., the ability of the test to correctly identify people without the measured process). Large multicenter study with 286 paired
serum/CSF NFL measurements in MS patients demonstrated that CSF NFL (cNFL) has higher accuracy for predicting CEL or MS relapse (i.e., 75% specificity and 67% sensitivity with area under receiver operator characteristic [AUROC] curve 0.77) compared to sNFL (i.e., 80% specificity and 45% sensitivity with AUROC 0.66) (2, 3).

Because published cohorts reported only modest correlations between cNFL and sNFL (i.e., explaining between 40-60% of variance (5-7)), we asked whether we could improve accuracy of sNFL in predicting MS activity (i.e., CEL or relapse) and MS severity (i.e., rate of MS progression) by identifying physiological confounding factors that affect sNFL concentrations. Because cNFL more accurately identified patients with MS activity than sNFL, we hypothesized that by adjusting sNFL measurements for physiological covariates that may affect release of NFL from CNS axons, distribution volume of sNFL and its metabolism using multiple linear regression models, we would derive reproducible equation that will better approximate sNFL to cNFL values measured in parallel and therefore also predict MS activity with enhanced accuracy. Although we could not identify publications that assessed accuracy of NFL in predicting MS severity, we expected that cNFL will have superior accuracy than sNFL and that adjusting for same confounders will strengthen the accuracy of sNFL in predicting MS severity, too.
Results

*Adjusting sNFL levels for confounding factors using MLR model improves the correlation between measured and predicted cNFL*

We measured NFL levels in 1,138 matching CSF and serum samples collected from 571 subjects in 7 diagnostic groups: Healthy Donors (HD), Relapsing-Remitting MS (RRMS), Primary Progressive MS (PPMS), Secondary Progressive MS (SPMS), Clinically Isolated Syndrome (CIS), Non-Inflammatory Neurological Diseases (NIND), and Other Inflammatory Neurological Diseases (OIND). The NIND and OIND cohorts had evidence of central nervous system (CNS) injury. The cohort was split into training (2/3) and validation (1/3) dataset prior to running any analysis, balancing for the diagnosis. Because our first goal was to develop generally valid mathematical adjustment that would better approximate sNFL levels to cNFL, we deliberately included subjects with varied diagnoses.

The correlation between sNFL and cNFL in the training cohort showed that cNFL explains 57% of variance of sNFL (Figure 1A), which falls into higher-end estimate from the published studies (5-7). To search for confounding factors responsible for the remaining 43% of variance between cNFL and sNFL (Figure 1B), we assumed that NFL is released from CNS axons into CSF, which is then drained (partially via lymph) to blood. Because of prolonged dynamic of NFL release after acute injury, a steady state between cNFL and sNFL will ensue. We considered following confounding factors that may influence this steady state by modifying: 1. release of NFL from CNS (e.g., age); 2. distribution volume of sNFL (i.e., reflected by body mass index [BMI], height, weight, and estimated blood volume) and 3. sNFL metabolism/clearance from body (i.e., reflected by liver function tests: alanine transaminase [ALT], aspartate aminotransferase [AST], by phagocytosis via reticulo-endothelial system reflected by alkaline phosphatase [AP], by
protein metabolism reflected by blood urea nitrogen [BUN] and by kidney clearance reflected by serum creatinine and estimated glomerular filtration rate [eGFR]; Figure 1C). Of these, the stepwise multiple linear regression (MLR) retained only weight, AP, BUN, creatinine and age (Figure 1D and E).

Although our goal is to adjust sNFL measurements to better approximate cNFL, the mathematical strategy to derive such adjustment is counterintuitive: because we assume that source of sNFL is CNS (all subjects other than HD have CNS disease) and the route of NFL release is via CSF to blood, the sNFL concentration is cNFL modified by confounders. Without confounding factors, the sNFL would only depend on cNFL, which is approximated by linear equation derived from the training cohort in Figure 1F, where it explains 57% of variance. When applied to the independent validation cohort (Figure 1H) this linear model explains 53% of variance with very low p-value. In MLR model, the sNFL concentration depends on cNFL, but its concentration is further modified by 5 confounders, reflected by equation in Figure 1G derived from the training cohort, where it enhances proportion of variance explained from 57% to 67% (Figure 1F versus 1G; 10% absolute and 15% relative gain in accuracy). This gain in accuracy is reproducible, with improvement from 53% to 65% of variance explained in an independent validation cohort (Figure 1H versus 1I; 12% absolute and 18.5% relative gain in accuracy).

To assure that longitudinal samples (i.e., multiple samples per subject) do not affect the model performance, we show analogous validation in the cohort that contained only first collected sample per patient or median value of all samples per patient (Supplementary Figure 1).

Reshuffling the equation from the final MLR model, we now have correct adjustment of measured sNFL values that better predict cNFL:
sNFL-predicted cNFL = (Log10 sNFL - 0.005*Age + 0.004*Weight - 0.001*AP -
0.01*BUN - 0.14*Creatinine + 0.75) / 0.54

Adjusted sNFL correlates better with MRI CELs than measured sNFL

The most important question is whether the proposed mathematical adjustment enhances the clinical value of sNFL, such as its ability to predict MS activity.

Thus, for all subsequent studies, we will compare effects of measured cNFL, measured sNFL and sNFL-predicted cNFL (i.e., sNFL adjusted for validated confounders) on clinical and imaging outcomes in MS cohort only.

There are two approaches for assessing MS activity, one is to dichotomize patients into those who do, or do not have CELs or MS exacerbation at the time of NFL measurement and use the training cohort data to select optimal NFL value based on Area Under Receiver Operator Characteristic Curve (AUC; Figure 2A-F). This approach was used previously (2, 3) and is expanded here by assessing AUCs after applying NFL dichotomization cut-offs from the training cohort (measured cNFL = 3699 pg/ml, measured sNFL = 57 pg/ml, and sNFL-predicted cNFL = 5172 pg/ml; Figure 2A, 2B, and 2C) to the validation cohort. We reproduced published observation that cNFL is stronger predictor of MS activity as compared to sNFL in the training cohort: i.e., cNFL achieved AUC 78.4% vs 61.8% for sNFL (Figure 2A and 2B). This hierarchy was validated in the independent cohort, where the AUC of cNFL was 73.5% and sNFL 65.1% (Figure 2 D and 2E). The mathematical adjustment of sNFL for confounding factors increased AUCs as compared to measured sNFL in both training (Figure 2B versus 2C; from AUC 61.8% to 69.2%) and validation (Figure 2E versus 2F; from AUC 65.1% to 75.3%) cohorts.
However, simply dichotomizing patients is suboptimal. The true clinical value of laboratory tests resides in their quantitative aspect. For example, dichotomizing liver function tests into normal and abnormal would not inform clinical management of the patients in abnormal category. Therefore, we also assessed accuracy of NFL in estimating the level of disease activity, by generating models that predict the number of CELs in the independent validation cohort.

We observed that cNFL correlates stronger with number of CELs than sNFL, in linear regression models, cNFL explains 21% of variance of CELs (p < 2.2x10^{-16}; Figure 2G), while sNFL explains 5.6% of variance (p = 3.1x10^{-06}; Figure 2H) in the training cohort. Similar correlations were observed in an independent validation cohort (Figure 2J and 2K). Adjusting sNFL for confounders (Figure 2I and 2L) improved correlation with the number of CELs in comparison to measured sNFL by increasing the variance explained from 5.6% to 12% in the training cohort (Figure 2H versus 2I) and from 18% to 28% in an independent validation cohort (Figure 2K versus 2L); This translates into relative improvement of 36% in the validation cohort.

Supplementary Figure 2 contains additional sensitivity analyses assessing exponential models and Poisson regressions that show analogous results. We also tested if the improvement was statistically significant and observed significantly lower confounder-adjusted sNFL residuals of CEL from the Poisson regression models compared to unadjusted sNFL residuals (p = 0.003, paired Wilcoxon signed ranked test).

This validates the hypothesis that adjusting measured sNFL levels for identified covariates meaningfully improves its ability to predict MS activity in an independent validation cohort.
All NFL measurements are poor predictors of MS severity, but sNFL shows at least weak, but significant correlations with MS severity outcomes.

As increased levels of NFL reflect MS-related acute CNS injury, with cNFL demonstrating higher accuracy than sNFL, we hypothesized that elevated NFL levels will be also associated with faster accumulation of clinical disability and expected that cNFL will again demonstrate higher accuracy.

MS-related clinical disability is traditionally measured by ordinal Expanded Disability Status Scale (EDSS) (8). However, natural history studies showed that MS patients progress, on average, by approximately 1 EDSS point per decade. This has been validated in MS clinical trials, where approximately 10% of placebo-treated patients experience sustained disability progression on EDSS for each 1 year of trial duration. Thus, EDSS cannot correctly quantifying patient-specific MS progression in observational studies with follow-up shorter than 10 years, which represented most patients in our cohort. Consequently, we assessed correlations between three NFL values and cross-sectional MS severity outcomes: Multiple Sclerosis Disease Severity Scale (MS-DSS) (9), Multiple Sclerosis Severity Scale (MSSS) (10), and Age-Related Multiple Sclerosis Severity (ARMSS) (11). Note that these outcomes reflect past rates of disability accumulation by normalizing cross-sectional disability to patients age (ARMSS and MS-DSS) or disease duration (MSSS).

Contrary to our predictions, sNFL consistently outperformed cNFL (Figure 3). None of the severity outcomes showed statistically significant correlation with cNFL, but all three showed a weak but statistically significant correlation with sNFL (Figure 3A). Specifically, unadjusted sNFL measurements explained 2.6% (MS-DSS), 6.3% (MSSS), and 3.5% (ARMSS) of MS severity variance in the training cohort. These unexpected observations were validated in an
independent validation cohort, where sNFL explained 12% of MS-DSS, 13% of MSSS and 4.3% of ARMSS (Figure 3B). As would be expected from the adjustment that better approximates sNFL to cNFL levels, adjusting sNFL for covariates generally weakened correlations with MS severity outcomes, although all remained statistically significant (Figure 3A and 3B).

Why sNFL correlates stronger with MS severity outcomes compared to cNFL?

We generated and tested two mutually non-exclusive hypotheses that may explain why sNFL correlates stronger with MS severity compared to cNFL: 1) brain atrophy associated MS progression leads to dilution of cNFL due to compensatory increase in CSF volume (Figure 4A); and 2) Spinal cord (SC) injury, associated with MS disability, such as injury to lower motor neurons or autonomic nervous system, leads to release of NFL from axons of peripheral nerves into blood, bypassing the CSF (Figure 4D).

Both hypotheses were tested by focusing on subjects with either comparable sNFL levels and highly different cNFL levels (testing Hypothesis #1), or, conversely, comparable cNFL levels and highly different sNFL levels (testing Hypothesis #2). We achieved these using quartiles of appropriate sNFL-cNFL residuals (Figure 4B and E) or by propensity score-matched samples (Supplementary Figure 3). Because both approaches provided analogous results, we present here only the simpler approach.

To test the hypothesis that increased CSF volume dilutes cNFL concentration, we asked whether subject with proportionally lower cNFL compared to sNFL concentrations have higher brain atrophy (and thus enlarged CSF volume) compared to subjects with proportionally higher cNFL concentration. We tested two brain atrophy outcomes (Figure 4A): 1) a fully quantitative brain
parenchymal fraction (BPFr) measured retrospectively and 2) prospectively-acquired
semiquantitative grading of the atrophy into four categories (none, mild, moderate, and severe)
as part of previously validated Combinatorial MRI scale (COMRIS) of CNS tissue destruction (12).

Consistent with stated hypothesis, people with proportionally lower cNFL concentration had
marginally significant increase in brain atrophy measured by semi-quantitative outcome
(p=0.0049) and BPFr (p=0.053) in the training cohort. However, these weak differences were not
validated in an independent validation cohort (Figure 4C).

To test the second hypothesis, we used prospectively acquired SC injury outcomes (Figure 4D).
First outcome is a semi-quantitative assessment of lesion load and atrophy of the upper cervical
SC graded from brain MRI images that extend to C5 level. This outcome has been previously
validated as clinically meaningful (12, 13). Because it does not capture the damage to the
thoracic or lumbosacral spinal cord, we employed complementary information from the
NeurEx™, which provides granular measurements of neurological disability (14). Two parts of
NeurEx™ can be used for our purpose: grading of muscle atrophy, as surrogate of injury to
lower motor neurons and, by inference, to associated motor axons of peripheral nerves, and
bowel, bladder, sexual, and autonomic (BBSA) dysfunctions, which in MS is likely caused by
injury of the autonomic neurons that project axons to the autonomic ganglia.

Using these two outcomes we observed statistically significant increase in both imaging and
clinical SC damage outcomes in the subgroup of subjects with proportionally higher sNFL
compared to cNFL, fitting the proposed model. These findings were robustly validated in an
independent validation cohort (Figure 4F and Supplementary Figure 4).
We conclude MS-associated SC injury leads to release of NFL to blood, bypassing the CSF, likely because of Wallerian degeneration of peripheral axons. Thus, SC damage, which is a strong predictor of MS severity, is preferentially reflected by sNFL as compared to cNFL, leading to stronger correlation of sNFL with MS severity outcomes. This conclusion is supported by the fact that including SC damage outcomes (i.e., MRI SC atrophy, muscle atrophy and BBSA) into the MLR model further strengthens the correlation between measured and predicted sNFL in an independent cohort (Supplementary Figure 5).
Discussion

We started this work with the premise that cNFL is clinically more relevant biomarker of CNS injury than sNFL and that we might enhance the clinical value of sNFL by adjusting for relevant confounders.

We validated this premise partially: we identified general confounders and validated a MLR model that adjusted sNFL to better approximate cNFL concentrations. The covariates selected by MLR model are logical and affect sNFL concentration in biologically predictable manner: high BUN, AP and creatinine increase sNFL concentrations as they slow down NFL metabolism and clearance, while diluting NFL in larger distribution volume reflected by higher weight lowers sNFL concentration.

Some of these confounders (e.g., age, BMI and Cr) were inferred in previous studies analyzing only sNFL in HVs (15)(16) or in patients with diabetes mellitus and renal dysfunction (16). However, lack of matched cNFL concentrations in these studies precludes eliminating the alternative explanation that these confounders, such as renal dysfunction, cause subclinical axonal damage, which increases sNFL. Current study unequivocally rules out this alternative explanation and extends previous studies by identifying most comprehensive set of confounders and showing that this comprehensive adjustment of sNFL values further improves their correlation with cNFL. Most importantly, when we applied these adjustments to MS cohort only, we observed that covariate-adjusted sNFL demonstrated statistically significant increase in correlations with number of CELs and improved dichotomized prediction of MS activity when compared to measured sNFL. These improvements were validated in an independent validation cohort, where it reached 36% relative enhancement of accuracy to predict presence of CEL. This is substantial improvement for a test aimed to be used at patient level. These data validate and
strengthen previous reports of using sNFL to identify patients with MS activity; in comparison to previously reported AUC = 0.66 for sNFL (measured in the training cohort only) (2, 3), we achieved AUC = 0.753 in an independent validation cohort for covariate-adjusted sNFL. Although this AUC approaches accuracy of clinically meaningful cross-sectional test, it should be noted that the test has high specificity (96.7%) but poor sensitivity (29.2%). Therefore, positive test adds clinical value in identifying MS patients with disease activity, e.g., during monitoring of treatment efficacy. However, due to poor sensitivity, a clinician must supplement NFL measurements with CNS imaging, perhaps performed less frequently, to verify that prescribed treatment truly abrogated formation of new MS lesions.

The novel and intriguing findings from our study is the demonstration that sNFL outperforms cNFL in correlating with cross-sectional MS severity outcomes and that this is due to the ability of sNFL to capture NFL released from two different sources: axons contained in the CNS, but also from spinal cord injury affecting lower motor neurons and autonomic nervous system. The strength of our study resides in prospective acquisition of complementary clinical and imaging data that reliably capture SC injury and associated damage to lower motoric and autonomic neurons. These data were locked into database before any NFL measurements were collected, and NFL was measured in the blinded fashion by investigators who had no access to clinical or imaging data. The congruency of observations between training and validation cohorts and very low p-values provide a high confidence that our conclusions are valid. The fact that adding imaging and clinical outcomes of SC damage to the MLR model further strengthens correlation between cNFL and sNFL shows that release of NFL from these SC sources is captured only by sNFL, but not by cNFL. Although we logically infer that the source of NFL associated with imaging and clinical outcomes of SC injury is Wallerian degeneration of peripheral axons, we
have not measured the peripheral nerve injury in this study. Nevertheless, our conclusion, novel for MS field, is supported by observation that subjects with PNS disorders, such as Guillain-Barre Syndrome (GBS), have proportionally higher sNFL to cNFL levels compared to subjects with CNS disorders such as Alzheimer’s disease, or compared to patients with combined involvement of CNS and peripheral axons such as Amyotrophic lateral sclerosis (17).

Despite its ability to (at least partially) reflect both brain and SC injury, the power of sNFL measurement to predict MS severity is, unfortunately, rather weak (i.e., explaining between 2.6-13% of variance of different MS severity outcomes). This observation suggests that sNFL will have limited value in guiding therapeutic decisions in patients who no longer form CELs. However, our dataset is suboptimal for determining prognostic value of sNFL in MS: this is an observational, natural history cohort with duration of follow-up less than 10 years in most patients, which precluded us from evaluating ability of sNFL to predict (future) rates of EDSS progression. Recent study evaluated the ability of sNFL to predict disability progression in two placebo-controlled Phase 3 clinical trials of progressive MS: EXPAND trial that evaluated efficacy of siponimod in SPMS and INFORMS trial that evaluated efficacy of fingolimod in PPMS (4). This study reported that sNFL (the authors measured plasma NFL, which is comparable to sNFL) dichotomized to “low” (< than 30 pg/ml) and “high” (≥30 pg/ml) has significant predictive value to identify patients at risk of MS progression. Unfortunately, this study did not report AUC, sensitivity/specificity and predictive values of dichotomized sNFL, which are necessary to assess clinical utility of a test. Patients with CELs constituted 21.6% of EXPAND and 10.1% of INFORMS participants and in both trials, presence of CELs was the strongest predictor of elevated sNFL. Nevertheless, the subgroup analyses demonstrated that even patients without CELs had statistically significant increase in the risk of disability.
progression if they had high sNFL. But this study analyzed thousands of sNFL samples and achieved marginal p-values for subjects without CELs (e.g., p=0.0274 for n=1147 predicting 3 months confirmed disability progression in EXPAND trial). Because both effect size and number of subjects contribute to p-value, the marginal p-value can be obtained from such large cohorts only if the effect size is very low. Thus, these data, generated from the gold-standard clinical trials support our conclusion that while sNFL correlates with MS severity on a group level, its accuracy is likely too low to be clinically meaningful on a patient level. We discourage clinicians to interpret ours, or (concurring) group data from other publications as proof that sNFL provides actionable insight on a patient level when managing people with MS who no longer form CELs.

Last question is why doesn’t sNFL predict MS progression with much higher accuracy when it is likely that neuronal loss causes development of brain atrophy and MS disability? Loss of neurons must lead to loss of axons, so why so many people with progressive MS have sNFL levels indistinguishable from HDs? We speculate that the answer lies in the spatial-temporal dynamics of axonal damage. Pathology studies show concentrated axonal transections inside acute MS lesions, but not outside of them. Such temporary and spatially concentrated axonal transections must release large amounts of NFL that likely overwhelms local phagocytes, leading to spill-over of NFL into circulation. But if neuronal loss is distributed over large CNS volume, microglia may have time to phagocytose dissolving axons and the NFL spillover is minimized. The type of neuronal death is probably also important: inflammation causes axonal transections associated with NFL release, whereas orderly apoptosis retains neuronal molecules encapsulated by membranes before they are phagocytosed.

Obviously, these are only speculations, difficult to investigate in living systems. Nevertheless, ours and thus far all reported NFL studies in MS demonstrate that non-physiological increase in
NFL preferentially reflect acute axonal injury associated with formation of focal MS lesions.

NFL levels are rather insensitive for capturing slow neuro-axonal loss associated with MS progression in people who no longer form CELs. If NFL should play a role in clinical management of such MS patients, we need publications that quantify predictive value of NFL on a patient level.
Methods

Subjects

Matching CSF and serum samples (1,138 each) were prospectively collected from 571 subjects (Table 1) from 7 diagnostic categories: HD, RRMS, PPMS, SPMS, CIS, NIND, and OIND. Collected samples were split into training (2/3) and validation (1/3) cohorts, controlling for diagnoses, and keeping longitudinal samples in the same cohort to assure complete independence of the subjects in two cohorts.

All laboratory, clinical, and MRI outcomes (see Supplementary Methods) were prospectively collected into database, quality controlled during weekly clinical care meetings and locked to prevent further modifications. These data were collected prior to blinded evaluation of CSF and serum NFL levels from cryopreserved samples.

Laboratory, clinical, and MRI outcomes

Body measurements were taken, and laboratory tests were performed at the time of CSF/blood collection at the NIH Department of Laboratory Medicine and recorded in the NIH electronical medical records. MS patients underwent a full neurological exam and brain MRI at the time of sample collection. The neurological exam was documented electronically using NeurEx™ App (14) that contains algorithms calculating traditional disability scales (e.g. EDSS, including Kurtzke functional system scores) that eliminate noise stemming from inaccuracy of translating neurological examination into disability scales by clinicians. The research brain MRI (with or without gadolinium contrast) was performed on 1.5T and 3T scanners. MRI sequences included T1-magnetization-prepared rapid gradient-echo (MPRAGE) or fast spoiled gradient echo.
(FSPGR) and T2-weighted three-dimensional fluid attenuation inversion recovery (3D FLAIR) sequences that were reviewed and graded by a board-certified neurologist and recorded using previously published Combinatorial MRI Scale of CNS tissue destruction (COMRIS) tool (12) into research database. The brain MRI protocol used extends sagittal and axial cuts distally to C5 level, allowing determination of semi-quantitative (semi-qMRI) MRI biomarkers of medulla/upper SC atrophy and lesion load. The quantitative MRI outcomes (e.g., brain parenchymal fraction) were generated using cloud-based medical image-processing platform, QMENTA, using LesionTOADS algorithm (18). MS severity outcomes: MS-DSS, MSSS, and ARMSS were calculated as described (9-11). While MSSS and ARMSS are both based on EDSS related to disease duration and age, respectively, MS-DSS is a more complex, machine learning-based model with the strongest variable being Combinatorial weight-adjusted disability score (CombiWISE (19))/Age.

Sample collection

Samples were collected following the laboratory Standard Operating Procedures. Briefly, CSF, collected by lumbar puncture, was kept on ice and processed within 15 min of collection by centrifugation at 335g for 10 min at 4°C; the supernatant was aliquoted and stored at -80°C. Blood was collected by venipuncture (SST tube), incubated at room temperature for 30 min, spun at 2,000g for 10 min at 4°C. The aliquoted serum was stored at -80°C. Personnel processing samples were blinded to patients’ diagnoses/clinical/MRI outcomes.

cNFL ELISA
As NFL concentration in CSF is higher (~10-to-100-fold higher than sNfL), it can be reliably measured with a comparatively less sensitive assay, such as ELISA. We measured cNFL concentrations using solid-phase sandwich ELISA (UmanDiagnostics, Umea, Sweden; Catalog number: 10-7002 RUO; Lower Limit of Detection (LLoD): 33 pg/ml; see Supplementary Methods).

All samples were diluted 1:2 with provided sample diluent and then analyzed blindly. Samples were analyzed on multiple plates; location of samples on each plate was randomized and a control sample was analyzed on each plate. The coefficient of variance (CV) for the control sample across all plates was 6.6%, confirming the assay precision and reproducibility.

**sNFL SimoaTM Assay**

Serum NFL levels were analyzed using a Simoa™ assay kit (Quanterix, Billerica, MA, USA; Product number: 103186; LLoD: 0.038 pg/ml) on Simoa HD-1 analyzer (see Supplementary Methods).

All samples were diluted 1:4 with provided sample diluent using on board dilution functionality, and then analyzed blindly. Samples were analyzed in two batches (batch 1: 12 plates and batch 2: 4 plates); each plate contained two quality control (QC) samples provided with kit, one for low (C1) and one for high (C2) concentration. The CVs for measured concentrations of QC samples were within the acceptable range (batch 1: C1=9.8%, C2=9.8%; batch 2: C1=9.0%, C2=7.7%), confirming the assay precision.

Though we have used two different assays, ELISA and SIMOA, to assess CSF and serum respectively, both assays use identical antibodies. ELISA is cheaper and has sufficient sensitivity
for measuring cNFL, while enhanced sensitivity of SIMOA assay is necessary to reliably measure sNFL. To assure that no bias was introduced by using different assay for NFL measurements, we measured 68 CSF samples by both assays and confirmed that they yield identical results ($R^2 = 0.97$, $p<0.001$; Supplementary Figure 7).

Statistics

All modeling/analyses/plots were performed in R Studio Version 1.1.463 (R version 4.0.2) (20). Simple and multiple linear regression (MLR) models were generated using `lm` function (“stats” package (20)). Correlations between variables were assessed using `stat_cor` function (“ggpubr” package (21)), generating Pearson correlation coefficient ($r$), Coefficient of variance ($R^2$) and p-value. Concordance correlation coefficient (CCC) was calculated using `epi.CCC` function of the “epiR” package. Final multiple linear regression model was selected using stepwise algorithm in `stepAIC` function (“MASS” package(22)). Differences between groups were evaluated by `stat_compare_means` function (“ggpubr” package (21)) using unpaired two-sided `wilcox.test` or `t.test` method.

Prediction models of MRI CEL were developed using logistic regression (glm function of the “stat” package) (20) Optimal cutoff for the predictive models was calculated using the `optimalCutoff` function of the “InformationValue” package (https://cran.r-project.org/web/packages/InformationValue). The area under the receiver operating characteristic curve (AUROC) was calculated using the `roc` function of the “pROC” package (23) and the specificity and sensitivity were calculated using specificity and sensitivity functions of the “caret” package (https://cran.r-project.org/web/packages/caret). The NFL cutoffs depicted
in Figure 2A-C were calculated as mean of the highest NFL value below the optimal model cutoff and the lowest NFL value above the optimal model cutoff.

To test whether brain atrophy can explain superiority of sNFL over cNFL we generated NFL residuals by subtracting the variance of cNFL explained by sNFL. Then we calculated quartiles of the NFL residual and removed samples falling within the interquartile range (IQR). Samples with NFL residuals below the first quartile represented patients with measured cNFL much lower than what would be predicted by the simple linear regression model. To test whether spinal cord damage could explain superiority of sNFL in predicting MS severity, we generated NFL residuals, by subtracting variance of the sNFL explained by the measured cNFL. Then we eliminated samples with NFL residuals within the IQR, resulting in a group of samples with measured sNFL higher than what would be predicted by the model and samples with measured sNFL levels that were lower than what the model predicted. Differences between the samples from the first and the third quartile were evaluated using unpaired wilcox.test or t.test method.

Propensity score matching was performed using matchit function with “full” method (“MatchIt” package(24)). Differences between propensity score-matched groups were evaluated by stat_compare_means function (“ggpubr” package (21)) using paired wilcox.test or t.test method.

Poisson regression models were generated using glm function.

Although we provide raw p-values that have not been adjusted for multiple comparisons, all p-values in the independent validation cohort would remain significant after the most conservative Bonferroni adjustment.

The raw data and R code will be provided as Supplementary information upon acceptance of the article.
**Study approval**

All samples were collected as part of a Natural History protocol “Comprehensive Multimodal Analysis of Neuroimmunological Diseases of the Central Nervous System” (ClinicalTrials.gov Identifier: NCT00794352) or as part of the “NIB Repository Protocol” (10-N-0210). The protocols were approved by the National Institutes of Health - Institutional Review Board (NIH-IRB). All participants signed a written informed consent.
Authors’ contributions


Acknowledgments

We would like to thank clinicians Alison Wichman and Jamie Cherup, research nurses Tiffany Hauser, Naomi Gathua, Jenifer Dwyer, and patient care coordinators Michelle Woodland, Kewonie Pumphrey, for their excellent patient care and scheduling work. We would like to thank laboratory technician Elena Romm for processing patient samples. Finally, we would like to thank our patients, their caregivers, and healthy volunteers for being partners in research – without them this work would be possible.

This study was funded by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health.
References


Figure 1: Variance between sNFL and cNFL concentrations. (A) Linear regression model between log10-transformed concentration (pg/ml) of sNFL and cNFL in the training cohort of samples where cNFL levels explain 57% of variance of sNFL levels. (B) Remaining 43% of variance shown as NFL residuals generated as differences between measured sNFL concentration and predicted sNFL concentration calculated from measured cNFL using linear regression model. (C) Eleven potential confounders related to distribution volume (BMI = body mass index, Est Blood Vol = estimated blood volume, Height, and Weight) protein metabolism/clearance (ALT = alanine transaminase, AP = alkaline phosphatase, AST = aspartate transaminase, BUN = blood urea nitrogen, Creatinine, and eGFR = estimated glomerular filtration rate), and Age were used as explanatory variables in a multiple linear regression model resulting in various importance represented as a t statistics of each variable in the model (D). Stepwise regression resulted in retention of 5 confounders in the model (E) that showed increased correlation between measured and predicted sNFL levels both in the training (G) and the validation (I) cohort in comparison to correlations between measured and predicted values using a simple linear regression model in the same training (F) and validation cohort (H). Confounders in bold are the ones selected in the multiple linear regression model that underwent stepwise regression. ns – number of samples measured, np – number of patients represented by the samples. Green line represents linear regression model with gray shading corresponding to 95% confidence interval.
**Figure 2**: Adjustment for five confounders improves correlation of sNFL with number of MRI contrast enhancing lesions (CEls) and eliminates noise. CEls have been used as surrogate outcome of blood brain barrier opening and active inflammation in the brains of MS patients. Logistic regression that predicts probability of CEL presence/absence and linear regression between NFL and total number of CEls have been tested.

On the left, a binomial regression classifier was generated to predict dichotomous outcome of present/absent CEL. The Area under the curve (AUC), sensitivity and specificity have been
calculated for classifiers using measured cNFL (A and D), measured sNFL (B and E), and sNFL-predicted cNFL (C and F) to predict probability of presence of CEL. Dotted line represents the best probability cut-off value determined in the training cohort with corresponding NFL concentration displayed above the line. Two-sided Wilcoxon two-sample test evaluated the significance of differences between two groups of patients.

On the right, linear model between number of CELs (y-axes - transformed as natural logarithm of [CEL+1]) and NFL (x-axes) shows higher predictive power of cNFL in both training (G) and validation (J) cohorts, compared to sNFL in training (H) and validation (K) cohorts. Adjustment of sNFL for five confounders (age, weight, alkaline phosphatase, blood urea nitrogen, and creatinine) improved the correlation with number of CELs in both training (I) and validation (L) cohorts compared to measured sNFL. Purple line represents linear regression model with gray shading corresponding to 95% confidence interval.
Figure 3: sNFL correlates better with MS disease severity outcomes than cNFL. Disease severity in MS is a measure of how fast patients accumulate disability; slow accumulation of disability over time results in low MS severity (green), fast accumulation of disability results in high MS severity (red). Because it is very hard to measure rates of disability progression prospectively and longitudinally, MS severity outcomes are collected cross-sectionally, measuring past rates of disability progression by normalizing disability to patient’s age (Age-Related Multiple Sclerosis Severity Score [ARMSS] and Multiple Sclerosis Disease Severity Scale [MS-DSS]) or disease duration (MSSS).

Correlation analysis of three MS severity outcomes: Multiple Sclerosis Disease Severity Scale (MS-DSS), Multiple Sclerosis Severity Score (MSSS), and Age-Related Multiple Sclerosis...
Severity Score (ARMSS) with 3 NFL values: measured cNFL, measured sNFL, and sNFL-predicted cNFL, in 2 two independent cohorts: training cohort (A) and validation cohort (B). Purple line represents linear regression model with gray shading corresponding to 95% confidence interval. Difference in number of patients/samples used for these analyses is because of exclusion of samples due to missing respective MS severity data.
Figure 4: Two hypotheses explaining superiority of sNFL in predicting MS severity. (A) Hypothesis 1: Dilution of cNFL due to brain atrophy while sNFL concentration remains unaffected. Brain atrophy was evaluated by brain parenchymal fraction (BPFr) and by semi-quantitative measure of brain atrophy (none, mild, moderate, and severe). (B) NFL residuals that fall within interquartile range (IQR, grey) were removed resulting in subset of samples with proportionally higher (above the 3rd quartile [teal]) and lower cNFL (below the 1st quartile [salmon]), with comparable sNFL levels. (C) Paired Wilcoxon Ranked Sum Test showed marginally significant difference in BPFr (top left) and total brain atrophy (bottom left) between samples with different cNFL levels in the training cohort. These observations were not confirmed in the validation cohort (top and bottom right).

(D) Hypothesis 2: Increase of sNFL due to spinal cord (SC) damage. NFL from damaged peripheral nerves and SC roots is released directly into blood, increasing sNFL concentration while cNFL remains unchanged. SC damage was evaluated using a semi-quantitative MRI outcome (a sum of lesion load and atrophy at the level of medulla and cervical spine) and by clinical outcome capturing damage of lower motor neurons (sum of muscle atrophy scores) and damage to peripheral/autonomic nervous system (score for bowel, bladder, sexual, and autonomic dysfunctions) generated from neurological exam digitalized using NeurEx™ App.

(E) NFL residuals that fall within IQR (grey) were removed, resulting in subset of samples with proportionally higher sNFL (above the 3rd quartile [teal]) and lower sNFL (below the 1st quartile [salmon]), with comparable cNFL levels. (F) Paired Wilcoxon Ranked Sum Test showed statistically significant difference in MRI (top left) and clinical (bottom left) outcomes between samples with different sNFL levels in the training cohort; the observed differences were confirmed in the validation cohort (top and bottom right).
### Table 1: Demographic details of training and validation cohorts.

<table>
<thead>
<tr>
<th></th>
<th>Training cohort (n=376)</th>
<th>Validation cohort (n=195)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong></td>
<td><strong>CIS</strong></td>
<td><strong>HD</strong></td>
</tr>
<tr>
<td><strong>N (Female/Male)</strong></td>
<td>10/4</td>
<td>16/11</td>
</tr>
<tr>
<td><em><em>Avg age in years</em> (SD)</em>*</td>
<td>41.0 (13.0)</td>
<td>37.9 (13.6)</td>
</tr>
<tr>
<td><strong>Age range</strong></td>
<td>20.8-64.2</td>
<td>19.7-71.3</td>
</tr>
<tr>
<td><strong>N of matching CSF - serum samples</strong></td>
<td>18</td>
<td>42</td>
</tr>
<tr>
<td><strong>Avg N of samples per subject (SD)</strong></td>
<td>1.3 (0.6)</td>
<td>1.6 (0.9)</td>
</tr>
<tr>
<td><strong>Avg cNFL in pg/ml (SD)</strong></td>
<td>1053 (1150)</td>
<td>544 (230)</td>
</tr>
<tr>
<td><strong>Avg sNfL in pg/ml (SD)</strong></td>
<td>10.1 (6.2)</td>
<td>8.7 (5.2)</td>
</tr>
</tbody>
</table>
*: Age at the first sample collection for respective subjects, N: number, Avg: average, and SD: standard deviation.