Multicenter immunoassay validation of cerebrospinal fluid neurofilament light: a biomarker for neurodegeneration

Authors: Miller A-M, Rutkowska A, Bahl JM, Herukka S-K, Koel-Simmelink MJA, Kruse N, Mollenhauer B, Siloaho M, Skinningsrud A, ZETTERBERG H, Teunissen CE, Lawlor BA

Key words: biomarker, validation, neurofilament light (NfL), cerebrospinal fluid (CSF), multi-centre

Corresponding Author:

Anne-Marie Miller
Mercer's Institute for Research on Aging
St. James's Hospital
Dublin 8
Ireland.
Ph: 00 353 1 4284531
Email: milleram@tcd.ie

Contents

Abstract

Neurofilament light (NfL), a putative cerebrospinal fluid (CSF) biomarker, has proven capacity to support the diagnosis of, or differentiate between, various sub-types of neurodegenerative disorders. It also functions as a marker of disease severity and prognosis with potential to contribute to patient stratification in clinical and scientific research. Here, a commercially available NfL enzyme-linked immunosorbent assay (ELISA) (UmanDiagnostics, Umeå, Sweden) was validated in a multi-centred setting using standardised operating procedures (SOPs) devised as part of the Joint Program for Neurodegenerative Diseases (JPND) BIOMARKAPD project. Validation parameters included (i) sensitivity; (ii) precision (intra- and inter-assay); (iii) recovery (selectivity); (iv) linearity of dilution; (v) parallelism and (vi) inter-laboratory precision. The data showed good assay sensitivity and precision. Inter-laboratory precision was sub-optimal. Reasons for this have yet to be elucidated but may include pre-analytical methodological variation or differing levels of operator experience with the assay and / or SOPs. Over-all this assay is suitable for the quantification of NfL in human CSF but sources of inter-laboratory variation require further investigation, and need to be controlled for, before subsequent multi-centred validations of this assay be pursued and the true benefit of standardised practice in multi-site settings established.

Introduction

Fluid biomarkers, cerebrospinal fluid (CSF) biomarkers in particular, are increasingly being used to support the clinical diagnosis of Alzheimer's disease (AD) and their utility to support the differential diagnosis of other forms of dementia and neurodegeneration has also been acknowledged[1,2,3,4,5,6].

Neurofilament light (NfL) is one such candidate CSF biomarker. It comprises one of 4 cytoskeletal subunits of neuronal intermediate filaments and serves to maintain the shape of nerve cells and facilitate radial axonal growth hence maintaining neuronal conduction velocity[7,8]. Increased concentrations of NfL in the CSF are thought to reflect nerve cell damage or death[9] and have been shown to correlate with the severity of head trauma[10], to reflect the degree of degeneration in certain dementia subtypes[11] as well as increased rates of cognitive decline and disease progression in Alzheimer's patients[12]. Furthermore, NfL has demonstrated utility as a clinical indicator of disease severity, progression and prognosis as well as a support for patient stratification or a proxy measure of therapeutic efficacy in other neurodegenerative conditions such as multiple sclerosis (MS)[13], amyotrophic lateral sclerosis (ALS)[14] and Parkinson's disease (PD)[11,15-20].

CSF neurofilaments, similar to other CSF biomarkers, can be measured for clinical or research purposes using routine analytical platforms such as enzyme-linked immunosorbent assays (ELISAs) [18,20]. However, for medical practitioners and researchers to recognise and accept the use of neurofilaments, and other neurodegenerative CSF biomarkers into routine diagnostic processes or clinical and research programmes, the reliability and reproducibility of their quantification needs to be established[21-23]. This is only possible by stringently validating the performance parameters of the various analytical platforms prior to using them for routine clinical or research purposes. To date attempts at investigating the performance of assays for amyloid and neurodegenerative biomarkers of AD have proven to be problematic with unacceptable levels of inter-laboratory variation being reported[24,25]. More recently, a multicentre NfL ELISA assay validation experiment, which examined whether parameters such as pipette calibration, duration of colorimetric development or delay in analysis could contribute to the observed high levels of inter-laboratory variation reported that disparity in the accuracy of standard preparation, contributed most significantly to the observed large inter-laboratory coefficient of variation (59%) in that particular experiment[26].

The purpose of the present work was to implement a standardised immunoassay validation protocol across multiple international research centres and in doing so determine the suitability of an established ELISA assay to quantify CSF NfL concentrations to facilitate implementation in clinical practise. Novel biomarker validation experiments comprise a subtask of the BIOMARKAPD project,

a Joint Programme Neurodegenerative Disease (JPND) initiative, which seeks to optimise the performance of fluid biomarkers for the diagnosis of AD and PD. The standard operating procedures (SOPs) used in these experiments were devised specifically for the BIOMARKAPD novel biomarker assay validation project. They were designed for the partial validation of commercially available novel biomarker assays. Their purpose was to allow for easy implementation of standardised assay validation protocols and appraisal of the performance of the assay without having to assess the kits performance in relation to robustness i.e. its ability to perform despite slight methodological variation[27].

Methods

Study design

The validation experiments were performed according to ISO/IEC 17025, ISO 5725-2, ISO 5725-4 criteria and the validation protocol was devised with reference to several supporting guidance publications[28-35]. Six laboratories located in Sweden, Denmark, Germany, Norway, Finland and Ireland took part in the study. The laboratories, with varying levels of familiarity with the assay ranging from first time to experienced users, received five UmanDiagnostics NfL ELISA kits (cat. #10-7001, Umea, Sweden) and 20 clinical CSF samples of known NfL concentration for inter-laboratory accuracy analysis. For all other validation experiments participating laboratories used locally obtained CSF samples. Each laboratory was provided with identical immunoassay validation protocol and data reporting templates. All laboratories were required to assess the UmanDiagnostics NfL ELISA kits performance with respect to six validation parameters which included an assessment of the kit's (i) sensitivity, (ii) precision (inter- and intra- assay), (iii) recovery (selectivity), (iv) linearity of dilution, (v) parallelism and (vi) inter-laboratory accuracy. Acceptable performance for each of the parameters measured was set at +/- 15% of the optimum performance (100%) or a coefficient of variation (%CV) below 15%[36].

ELISA

The NfL ELISA kits were used for quantitative determination of NfL in human CSF samples. ELISAs were performed according to the manufacturer's instructions and as described before[18,26].

In brief, standards (lyophilised bovine NfL spinal cord concentrate - TBC) were re-suspended in sample diluent to give a top concentration of 10,000pg/mL. The top standard was vortexed briefly and kept at room temperature (RT). The remaining standards were prepared from serial dilutions of the top standard and are summarised in **Table 1**. Initially plates were washed (x3) with wash buffer (300µL), they were then flipped over and wash buffer residues removed by tapping the plate against blotting paper. Next 100µL of standard or sample were added to each well according to the experimental templates and incubated for 1hr at RT (20-25^{0}C) with agitation (800rpm). Subsequently,

plates were washed (x3) as described above and 100µL of freshly diluted tracer (biotin labelled antiNfL mAb) was added to each well and incubated for 45min at RT (20-25^{0}C) with agitation (800rpm). Again, plates were washed (x3) and 100µL of newly diluted conjugate (streptavidin-HRP) was added to each well and incubated for 30min at RT (20-25^{0}C) with agitation (800rpm). Next, plates were washed (x3) and 100µL of TMB was added to each well and incubated for 15min at RT (20-25^{0}C) with agitation (800rpm). Finally, 50µL of stop solution was added to each well and the absorbance read at 450 nm.

Sensitivity

The sensitivity of the assay was determined by running 16 blank samples (sample diluents). Mean and standard deviations (SDs) of the blanks were calculated and the lower limit of detection (LLOD) was calculated based on a signal of 10 SDs above the mean of the blanks.

Precision

To assess the precision of the assay, five quality control (QC) samples with pre-defined low and high NfL content within the linear range of the standard curve were collected by each laboratory. The samples were divided into 24 aliquots and stored at -80^{o}C until needed. For the intra-assay precision 12 replicates of each QC sample were run on day one. For the inter-assay precision three replicates of the samples were run on days two through five. The mean, SDs and %CV were calculated for repeatability (inter-assay) and intermediate precision (intra-assay).

Recovery

Recovery (selectivity) was assessed over two days. Five CSF samples of known low NfL content were used in these experiments. On both days the samples (three samples day one; two samples day two) were diluted 1:1 as per assay requirements and divided into four aliquots. Three of the aliquots were spiked with calibrator stock solution to low, medium and high NfL concentrations on the linear range of the standard curve. Samples were spiked with 5% volume to 95% volume of the diluted sample. The fourth aliquot was spiked with an equal volume of the sample diluent (neat sample).

Recovery was calculated according to the following formula:

%Recovery = ((Measured concentration$_{\text{spiked samples}}$ – Measured concentration$_{\text{neat samples}}$) / Theoretical concentration$_{\text{spiked}}$) x 100.

Linearity of dilution

To assess the linearity of dilution three pre-selected undiluted CSF samples with known high concentrations of NfL were selected and spiked with calibrator stock solution to a concentration

beyond the upper limit of detection (ULOD) of the assay (10µg/mL). Serial dilutions of each spiked sample were carried out until the concentration was below the lower limit of detection (LLOD).

Linearity of dilution was calculated according to the following formula:

%Linearity = (Observed concentration x Dilution factor / (Previous observed concentration x Dilution factor in the dilution series)) x 100.

Parallelism

Parallelism was assessed by investigating whether an equivalent relationship exists between the detection of the native NfL content in the CSF samples and the calibrator protein in assay diluent. For this purpose four CSF samples with known high NfL concentrations were chosen and a fifth sample of sample diluent was prepared by spiking the diluent with calibrator stock solution to a concentration equivalent to one of the native samples. A minimum of three and maximum of five serial dilutions were then performed on the samples to reach a NfL concentration below the lower limit of quantification. A parallel relationship was defined by comparing the slope of the calibrator to the mean slopes of the four CSF samples at various dilution factors. Parallelism was calculated by taking the mean signal of the four CSF samples at each dilution factor and comparing it to the signal of the spiked sample diluent. The % parallelism was calculated according the following formula:

%Parallelism = (Slope(a)Samples / Slope(a)Calibrator) x 100.

Inter-laboratory accuracy samples

Twenty de-identified CSF samples (3-5mL volumes) of pre-determined NfL concentration (190-25,000 pg/mL) were provided by the Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden. The study from which the CSF samples were collected was approved by the local ethics committee at the University of Gothenburg. CSF samples were collected through lumbar puncture in the L3/L4 or L4/L5 interspace into polypropylene tubes, centrifuged (2200 x g; 10min; 20^{o}C) and stored in polypropylene tubes at -80^{o}C pending biochemical analysis. Following shipment to Dublin, Ireland (the project co-ordinating site) the samples were thawed and further aliquoted (250 µL) into 0.5mL tubes and redistributed among the participating laboratories. Inter-laboratory accuracy was calculated by comparing the %CV of each of the 20 CSF sample measured across the six participating laboratories.

Data analysis

All data analysis was performed by the project coordinating site to ensure uniformity of results. Standardised project reporting templates were used to calculate the results were applicable. Data was analysed using Excel (Microsoft, Version 2010) and GraphPad (GraphPad Software Inc, Version 5.0, La Jolla, CA, USA) was used to create figures.

Results

Sensitivity

The mean LLOD reported by the six laboratories based on the 16 blank samples was 57.5 pg/ml and ranged from 23.3 pg/ml to 100 pg/ml **(Table 2)**. The manufacturer's reported LLOD is 32 pg/ml.

Precision

The average intra-assay CV based on the calculated NfL concentrations was 3.7% and ranged from 1.7 to 7.1% **(Figure 1a)**. The average inter-assay CV for the low NfL content samples was 13.2% (range 4.0 to 21.1%), for medium NfL content samples 12.4% (5.9 to 18.1%) and for the high NfL content samples 15% (9.8 to 21.2%) resulting in an average inter-assay CV of 13.5% for all samples **(Figure 1b)**.

Recovery (selectivity)

The reported recovery percentages were, in most cases outside the pre-defined range of 85-115%. Only samples spiked to medium NfL content in Lab #1 and #2 and to high NfL content in Lab #2 were within the pre-defined range **(Table 3)**.

Linearity of dilution

The data showed that 30% of all dilutions did not fall within the pre-defined range of 85%-115% **(Figure 2)**.

Parallelism

Two of the 6 participating laboratories (Lab #2 and #5) reported parallelism within the pre-defined range of 85-115% **(Figure 3)**.

Inter-laboratory precision

The inter-laboratory accuracy of the measurements of NfL concentrations was assessed with 20 CSF patient samples distributed between six laboratories. The inter-laboratory CV for all samples was 27% (range 16% to 32%) and was outside the pre-defined level of 15% **(Figure 4)**. The data showed that samples with the highest NfL content had the lowest %CV **(Figure 4a)**.

The differences in the standard curve OD values reported by the centres do not explain the high inter-laboratory %CV of the samples **(Figure 4b)**. For instance, difference in the OD values of the standard curve between Lab #3 and Lab#5 is on average 29% but the average difference in the calculated NfL concentrations is 14%. However, in another example, the difference in the OD values of the standard curve between Lab #1 and Lab #5 was 7% but the average difference of calculated NfL concentrations was 25%.

In order to elucidate possible sources of variation which might account for the high inter-laboratory %CV the participating laboratories were asked to complete a questionnaire (adapted from Petzold et al., 2010) regarding the sample and kit handling conditions during the series of experiments.

In brief, variation was found in the sample storage conditions and time taken to incubate the standards and samples following reconstitution of the calibrator (15-60min), minor variations in the time allowed for the colour reaction and whether plates were protected from sunlight were also observed. No difference was observed with respect to ELISA kit storage conditions, the temperatures at which the assays were executed and plate agitation details. In two of six cases pipettes were calibrated prior to the validation experiments. The 20 inter-laboratory patient CSF samples used by 5 of the 6 laboratories underwent two shipments on dry ice with the exception of those used by the co-ordinating site. The assay data was calculated using the 4 parameter logistic (4PL) nonlinear regression model by all six laboratories **(Table 4)**.

Discussion

The data showed that the assay has good sensitivity with an average LLOD reported by the participating laboratories of 57.5 pg/ml. Assay precision, as assessed by measuring inter- and intra-assay CV was good at 13.2% and 3.7% respectively. Importantly however, it must be noted that that maximal spiking using assay calibrator, particularly in the recovery and linearity of dilution experiments was not, in most cases, achieved. This is due to the fact that the lyophilised calibrator is pre-diluted 1:1 by the manufacturer to allow for the 1:1 dilution of the samples in accordance with the NfL ELISA assay protocol thus, once re-suspended in diluent, the true concentration of the calibrator is in reality half of what is assumed. This pre-dilution of the calibrator was accounted for during calculations by the data coordinating site so that the data reflected 'true' calibrator concentrations.

Recovery (selectivity)

Recovery was assessed by measuring non-native NfL in the presence of endogenous CSF matrix, selectivity of the assay in the six laboratories was for the most part outside the pre-defined range of 85-115%. Usually poor recovery points to interference in sample measurement by components in the matrix, such as blood contamination, which may affect antibody binding[37] and is usually resolved by the addition of additional assay diluent to samples[27]. However, in this instance confounding elements such as the blood contamination status of the samples was unknown, this fact, in combination with the calibrator concentration miscalculation means that the recovery data must be interpreted with caution. However, if the assay were to display poor selectivity the observed values would tend to deviate in one direction, that is, either positive or negative. However, the results reported in this experiment showed that the data seemed to deviate in both directions suggesting that factors, other than those traditionally thought to affect selectivity, contributed to the poor recovery. Thus in the absence of a matrix effect it must be considered that recovery may have been hampered by the use of non-native (bovine) NfL calibrator protein to spike the CSF samples which might behave differently than native (human) NfL protein in the CSF[37].

Linearity of dilution

It is important to be able to demonstrate that if the concentration of a protein of interest is above the ULOQ of an assay that it can be diluted, without affecting accuracy of measurement, to a concentration within the working range of the assay[34]. Linearity of dilution was suboptimal in these experiments as 30% of dilutions were outside the pre-defined acceptance range. Similar to the recovery experiments poor linearity of dilution can imply that the sample matrix and assay diluent affect analyte detection differently[27] but, again, it is difficult to assess the validity of these results due to the assumed starting concentration of the calibrator being incorrect. Consequently only two- or three-fold dilutions of samples were made in these experiments before the LLOD was reached when normally four or five-fold dilutions would be compared. The 'hook effect' can also be discounted as

an influencing factor in this experiment. The 'hook effect' is observed when the analyte in question is in very high concentrations beyond the theoretical upper limit of detection of the assay and can result in artificially low concentrations of analyte being reported[38]. However this phenomenon was not observed in these experiments as, again, the working concentration of the calibrator used for spiking was significantly lower than originally assumed.

Parallelism

We examined parallelism comparing standard bovine NfL in assay diluent and endogenous NfL in CSF. Antibody affinity was investigated by comparing the slopes for the native NfL in CSF versus the calibrator protein in assay diluent. As two of six centres observed a parallel relationship between their samples and the calibrator sample we cannot conclude that the antibodies used in the assay have poor affinity.

A failure to observe parallelism could point to poor robustness in this particular experiment; however this can only be established by a more in-depth systematic examination of critical factors such as pre-analytical variables pertinent to this validation parameter[28]. Alternatively, failure to observe parallelism, and indeed some of the other validation parameters, could be in part due to some of the less practiced operators experiencing difficulty with interpretation or complexity of the plate plans and SOPs. It has been argued that as NfL has the tendency to re-aggregate following suspension[39-42] that delays in sample and standard incubation could account for the failure to satisfactorily measure NfL in many experiments[26]. The data show that between 15 and 60 minutes were taken, from the time standards were reconstituted, to incubate the standards and samples on the ELISA plates. The observed delay could be in part due to a lack of operator familiarity with the assay protocol or project SOPs. On foot of this observation we recommend that for validation experiments the operator use both vials of standard stock solution supplied in the kits to prepare the standard curve and spike samples as promptly as practicable. By using both vials during the course of the validation time to incubation of standards and samples would be significantly cut, diminishing the possibility of NfL protein aggregation and/or degradation contributing to variation in the performance of this assay across sites. Furthermore, future validation projects could be further optimised by adopting alternative assay validation SOPs such as those recently published by BIOMARKAPD[43] which are designed to allow for the inexperienced operator.

Inter-laboratory precision

The inter-laboratory CV at 27% is poor but cannot, as in other assay validation experiments be attributed to lot to lot variation in the kits[44] as all centres taking part worked with kits from the same lot. Nonetheless, despite the variation the data do point to the positive effects of implementation of a standardised protocol in that no one laboratory reported consistently high or low values and discrepancies seem to be spread out among the laboratories. In support of the positive effect of implementation of validation protocols a previous multi-centre study, albeit with greater numbers of

participating laboratories, which did not use standardised protocols in the validation of the UmanDiagnostics NfL ELISA reported a significantly higher inter-laboratory precision CV of 59%[26].

In agreement with our observations, Petzold and colleagues also observed variations in the concentration of standards and their associated optical densities (OD)[26]. They noted that non-systematic error due to variations in the preparation of standards could be responsible for up to two-fold differences in sample concentration, and identified differences in the accuracy and types of pipettes used to make the standards and dispense samples as possible sources of variation. Our own data indicates that four of the six centres reported not calibrating their pipettes prior to the validation experiments and declared using different pipettes for the preparation of the assay standards as to those used for pipetting samples. However, given the available data we are unable to definitively identify a reason for the inter-laboratory variation and further studies examining pre-analytical variation within the context of using standardised SOPs in a multi-centre context should be conducted.

Limitations and recommendations

A regularised approach to assay performance is required to ensure the validity of intra- and inter-laboratory data and the performance of multi-site clinical or research studies. This will ultimately lead to the optimisation of novel biomarkers, increase clinician confidence in the veracity of CSF biomarker results, resulting in their increased incorporation into routine clinical research and diagnostic processes. The data tentatively show that a standardised SOP can lend itself to greater control of the evaluation and reporting of the performance of immunochemical assays but due to the inconclusive results in some of the parameters examined in these experiments the outcome of this validation should be interpreted with caution. As yet too many factors, including pre-analytical procedures and the effect of delays in the execution of some of the experiments remain uncontrolled for in order to assess the true effect of standardised SOPs in a multi-centre study. In addition highlighting the dilution status of the lyophilised calibrator and use of both kit calibrators for any spiking experiments would greatly enhance the success of subsequent validations of this assay. Nonetheless given the data we conclude that the UmanDiagnostics ELISA kit for detection of human NfL is suitable for routine NfL quantification. However, we caution against comparing patient data sourced from different centres until the reasons for the high inter-laboratory CV are identified. We also recommend that going forward SOPs which minimise procedural complexity and allow for operator naïveté should be considered for future multi-centre validation studies.

Acknowledgements

This work is part of the BIOMARKAPD project within the EU Joint Programme for Neurodegenerative Disease Research (JPND). This project is supported through the following funding organizations under the aegis of JPND – www.jpnd.eu.

References

1. Blennow, K., de Leon, M. J., & Zetterberg, H. (2006). Alzheimer's disease. *Lancet, 368*(9533), 387-403.
2. Blennow, K., Zetterberg, H., Rinne, J. O., Salloway, S., Wei, J., Black, R., et al. (2012). Effect of immunotherapy with bapineuzumab on cerebrospinal fluid biomarker levels in patients with mild to moderate Alzheimer disease. *Arch Neurol, 69*(8), 1002-1010.
3. Burman, J., Zetterberg, H., Fransson, M., Loskog, A. S., Raininko, R., & Fagius, J. (2014). Assessing tissue damage in multiple sclerosis: a biomarker approach. *Acta Neurol Scand, 130*(2), 81-89.
4. Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol, 6*(8), 734-746.
5. Hall, S., Ohrfelt, A., Constantinescu, R., Andreasson, U., Surova, Y., Bostrom, F., et al. (2012). Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or parkinsonian disorders. *Arch Neurol, 69*(11), 1445-1452.
6. Lewczuk, P., & Wiltfang, J. (2008). Neurochemical dementia diagnostics: State of the art and research perspectives. *Proteomics, 8*(6), 1292-1301.
7. Lee, M. K., & Cleveland, D. W. (1996). Neuronal intermediate filaments. *Annu Rev Neurosci, 19*, 187-217.
8. Yuan, A., Rao Veerana, M.V., Nixon, R.A. (2012). Neurofilaments at a glance. J Cell Sci, 125(14), 3257-3263.
9. Landqvist Waldö, M., Frizell Santillo, A., Passant, U., Zetterberg, H., Rosengren, L., Nilsson, C., Englund, E. (2013). Cerebrospinal fluid neurofilament light chain protein levels in subtypes of frontotemporal dementia. *BMC Neurology*, *13*(54), 1-8.
10. Neselius, S., Brisby, H., Theodorsson, A., Blennow, K., Zetterberg, H., Marcusson, J. (2012) CSF-Biomarkers in Olympic Boxing: Diagnosis and Effects of Repetitive Head Trauma. *PLoS ONE*, *7*(4), e33606.
11. Scherling, C. S., Hall, T., Berisha, F., Klepac, K., Karydas, A., Coppola, G., et al. (2014). Cerebrospinal fluid neurofilament concentration reflects disease severity in frontotemporal degeneration. *Ann Neurol, 75*(1), 116-126.
12. Zetterberg, H., Skillbäck, T., Mattson, N., Trojanowski, J.Q., Portelius, E., Shaw, L.M., Weiner, M.W., Blennow, K. (2015) Alzheimer's Disease Neuroimaging Initiative. Association of Cerebrospinal fluid Neurofilament Light Concentration With Alzheimer Disease Progression. *JAMA Neurol*, *2*, 1-8.
13. Salzer, J., Svenningsson, A., Sundström, P. (2010) Neurofilament light as a prognostic marker in multiple sclerosis. *Mult Scler*, *16*(3), 287-92.
14. Tortelli, R., Ruggieri, M., Cortese, R., D'Errico, E., Capozzo, R., Leo, A., et al. (2012). Elevated cerebrospinal fluid neurofilament light levels in patients with amyotrophic lateral sclerosis: a possible marker of disease severity and progression. *Eur J Neurol, 19*, 1561-1567.
15. Abdo, W. F., Bloem, B. R., Van Geel, W. J., Esselink, R. A., & Verbeek, M. M. (2007). CSF neurofilament light chain and tau differentiate multiple system atrophy from Parkinson's disease. *Neurobiol Aging, 28*(5), 742-747.
16. Deisenhammer, F., Bartos, A., Egg, R., Gilhus, N. E., Giovannoni, G., Rauer, S., et al. (2006). Guidelines on routine cerebrospinal fluid analysis. Report from an EFNS task force. *Eur J Neurol, 13*(9), 913-922.
17. Lycke, J. N., Karlsson, J. E., Andersen, O., & Rosengren, L. E. (1998). Neurofilament protein in cerebrospinal fluid: a potential marker of activity in multiple sclerosis. *J Neurol Neurosurg Psychiatry, 64*(3), 402-404.
18. Norgren, N., Rosengren, L., & Stigbrand, T. (2003). Elevated neurofilament levels in neurological diseases. *Brain Res, 987*(1), 25-31.

19. Petzold, A. (2005). Neurofilament phosphoforms: surrogate markers for axonal injury, degeneration and loss. *J Neurol Sci, 233*(1-2), 183-198.
20. Rosengren, L. E., Karlsson, J. E., Karlsson, J. O., Persson, L. I., & Wikkelso, C. (1996). Patients with amyotrophic lateral sclerosis and other neurodegenerative diseases have increased levels of neurofilament protein in CSF. *J Neurochem, 67*(5), 2013-2018.
21. Mattsson, N., Zetterberg, H., Hansson, O., Andreason, N., Parnetti, L., Jonsson, M., et al. (2009). CSF Biomarkers and Incipient Alzheimer Disease in Patients with Mild Cognitive Impairment. *JAMA, 302*(4), 385-393.
22. Mattsson, N., Andreasson, U., Persson, S., Arai, S., Batish, S.D., Bernardini, S., et al. (2011). The Alzheimer Association external quality control program for cerebrospinal fluid biomarkers. *Alzheimers Dement, 7*, 386-395. Andreasson, U., Perret-Liaudet, A., van Waalwijk van Doorn, L. J., Blennow, K., Chiasserini, D., Engelborghs, S., et al. (2015). A Practical Guide to Immunoassay Method Validation. *Front Neurol, 6*(179), 179.
23. Carrillo, M.C., Blennow, K., Soares, H., Lewczuk, P., Mattsson, N., Oberoi, P., et al. (2013). Global standardization measurement of cerebral spinal fluid for Alzheimer's disease: An update from the Alzheimer's Association Global Biomarkers Consortium. *Alzheimers Dement, 9*, 137-140.
24. Lewczuk, P., Beck, G., Ganslandt, O., Esselmann, H., Deisenhammer, F., Regeniter, A., et al. (2006) International quality control survey of neurochemical dementia diagnosis. *Neurosci Lett, 409*, 1-4.
25. Verwey, N., Van der Flier, W.M., Blennow, K., Clark, C., Sokolow, S., De Deyn, P.P., et al. (2009). A worldwide multicentre comparison of assays for CSF biomarkers in Alzheimer disease. *Ann Clin Biochem, 46*(3), 235-240.
26. Petzold, A., Altintas, A., Andreoni, L., Bartos, A., Berthele, A., Blankenstein, M. A., et al. (2010). Neurofilament ELISA validation. *J Immunol Methods, 352*(1-2), 23-31.
27. Lee, J.W., Devanarayan V., Barrett, Y.C., Weiner, R., Allinson, J., Fountain, S., et al. (2006) Fit for purpose Method Development and Validation for Successful Biomarker Measurement. *Pharm Res, 23*(2), 312-328.
28. ISO/IEC17025. General Requirements for the Competence of Testing and Calibration Laboratories (2005).
29. ISO5725-2. Accuracy (Trueness and Precision) of Measurement Methods and Results– Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method (1994).
30. ISO5725-4:2003. Accuracy (Trueness and Precision) of Measurement Methods and Results - Part 4: Basic Methods for the Determination of the Trueness of a Standard Measurement Method (ISO 5725-4:1994).
31. Guidance for Industy - Bioanalytical Method Validation, Fed. Reg. 66:28526-28527 (2001).
32. Guideline on bioanalytical method validation, European Medicines Agency (2011).
33. Armbruster, D.A. and Pry, T. (2008) Limit of blank, limit of detection and limit of quantification. *Clin Biomchem Rev, 29* Suppl 1: S49-52.
34. Valentin, M.A., Ma, S., Zhao, A., Legay, F., Ayramesa, A. (2011). Validation of immunoassay for protein biomarkers: bioanalytical study plan implementation to support pre-clinical and clinical studies. *J Pharm Biomed Anal, 55*, 869-77.
35. DeSilva, B., Smith, W., Weiner, R., Kelley, M., Smolec, J., Lee, B., et al. (2003) Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res, 20*, 1885-900.
36. Guideline on validation of bioanalytical methods, European Medicines Agency (2009) Ref: EMEA/CHMP/EWP/192217/2009
37. Koel-Simmelink, M.J., Vennegoor, A., Killestein, J., Blankenstein, M.A., Norgren, N., Korth, C., Teunissen, C.E. (2014) The impact of pre-analytical variables on the stability of neurofilament proteins in CSF, determined by a novel validated SinglePlex Luminex assay and ELISA. *J Immunol Methods, 402*(1-2):43-9.
38. Dodig, S. (2009) Interferences in quantitative immunochemical methods. *Biochemia Medica, 19*(1), 50-62.

39. Liem, R.K. and Hutchinson, S.B. (1982) Purification of individual components of the neurofilament triplet: filament assembly from the 70 000-dalton subunit. *Biochemisty*, *21*(13), 3221.
40. Lee, M.K., Xu, Z., Wong, P.C. , Cleveland, D.W. (1993) Neurofilaments are obligate heteropolymers in vivo. *J Cell Biol*, *122*(6), 1337.
41. Carter, J., Gragerov, A., Konvicka, K., et al. (1998). Neurofilament (NF) assembly; divergent characteristics of human and rodent NF-L subunits. J Biol Chem, 273, 5101.
42. Perez-Olle, R., Leung, C.L., Liem, R.K. (2002). Effects of Charcot–Marie–Tooth linked mutations of the neurofilament light subunit on intermediate filament formation. J Cell Sci, 115, 4937.
43. Andreasson, U., Perret-Liaudet, A., vanWaalwijk vanDoorn, L.J.C., Blennow, K., Chiasserini, D., et al. (2015). A practical guide to immunoassay validation. *Front Neurol*, *6*(179)
44. Kruse, N. and Mollenhauer, B. (2015) Validation of a commercially available enzyme-linked immunoabsorbent assay for the quantification of human α-synuclein cerebrospinal fluid. *J Immunol Methods*, *426*, 70-75.

Figures

Standard Preparation:

Tube #	Sample Diluent	Standard from tube #	Concentration (pg/mL)
1	Per vial label		10000
2	300µL	300µL (1)	5000
3	300µL	300µL (2)	2500
4	360µL	240µL (3)	1000
5	300µL	300µL (4)	500
6	240µL	60µL (5)	100
7	300µL	0µL	0

Table 1: Standard dilutions per manufacturer's instruction

	Lab #1	Lab #2	Lab #3	Lab #4	Lab #5	Lab #6	Mean
LLOD (pg/ml)	40.8	23.3	56.4	100.0	75.0	49.6	**57.5**

Table 2. Lower limit of detection (LLOD). The mean LLOD reported by six laboratories was 57.5 pg/ml and ranged from 23.3 pg/ml to 100 pg/ml.

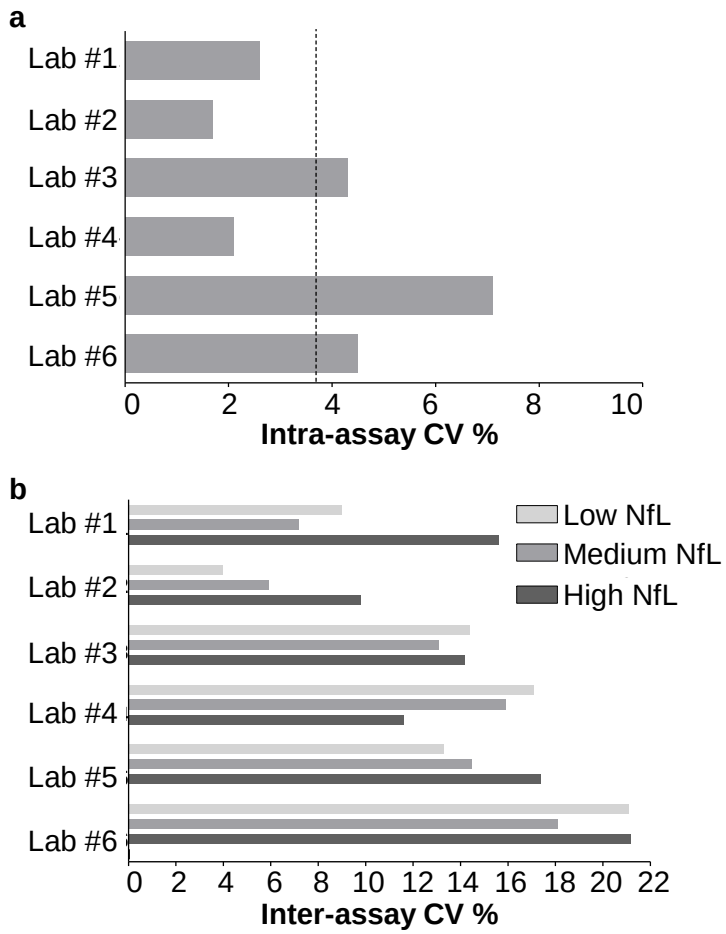


Figure 1. The assay had low inter- and intra- assay CV%. a) The average intra-assay CV% based on calculated NfL concentrations was 3.7% (range 1.7 to 7.1%). **(b)** The average inter-assay CV% for the low NfL content samples was 13.2% (range 4.0 to 21.1%), for medium NfL content samples 12.4% (5.9 to 18.1%) and for the high NfL content samples the average was 15% (9.8 to 21.2%) giving the average inter-assay CV for all samples among all the labs of 13.5%.

		Lab #1	Lab #2	Lab #4	Lab #5	Lab #6
Mean recovery						
Spike	**low**	207%	125%	129%	-96%	122%
	medium	115%	104%	133%	10%	128%
	high	132%	95%	139%	42%	128%
Recovery range (%)						
Spike	**low**	197-222	109-140	118-141	-454 - 129	93-109
	medium	58-157	95-116	118-148	-118 - 92	118-146
	high	119-152	93-100	125-151	-35 - 85	117-139

Table 3. The NfL Elisa assay used in the study had low selectivity and specificity. The selectivity of the assay was assessed using recovery experiments with three CSF samples with known low NfL content and spiked to low, medium and high NfL concentration with calibrator. The reported recovery percentages were, in most cases, either below or above the pre-defined range of 85-115%.

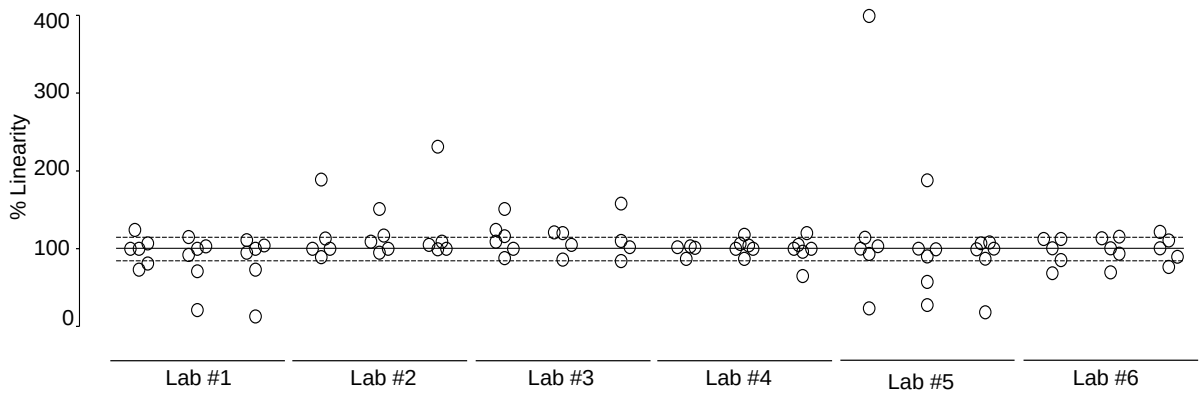
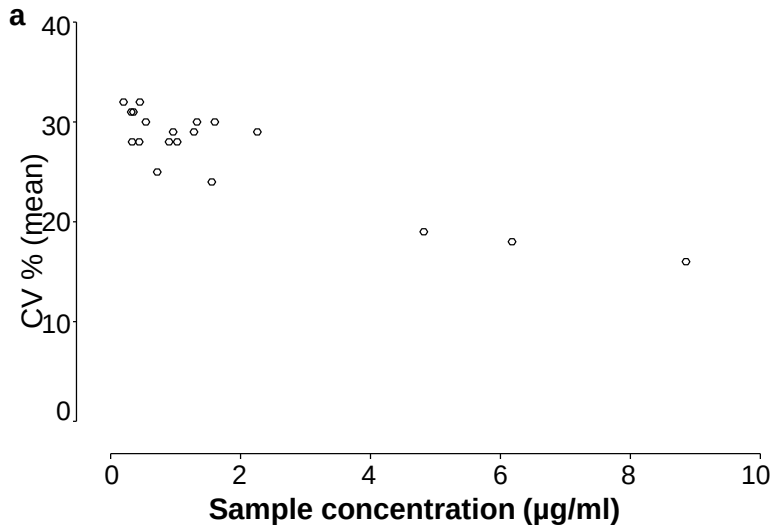


Figure 2. Linearity of dilution. The data showed that dilution of linearity was largely within the pre-defined range of 85%-115%. The graph shows % linearity for three samples per lab. The continuous line indicates 100% linearity and two dotted lines indicate lower (85%) and upper (115%) pre-defined ranges respectively. The "hook effect" was not observed in any of the laboratories (data not shown).

	Lab #1	Lab #2	Lab #3	Lab #4	Lab #5	Lab #6
Parallelism	144%	111%	174%	61%	109%	141%

Figure 3. Affinity was investigated looking at parallelism between the sample diluent spiked with NfL calibrator and four neat CSF samples. (a) Two out of five laboratories reported parallelism within the pre-defined range of 85 – 115% (Lab #2 and Lab #5). The remaining four laboratories reported parallelism outside the pre-defined range.



	Lab#2	Lab#3	Lab#4	Lab#5	Lab#6
Lab#1	51%/24%	31%/32%	24%/47%	7%/25%	24%/22%
Lab#2	-	30%/15%	38%/36%	50%/9%	36%/20%
Lab#3	-	-	17%/26%	29%/14%	15%/33%
Lab#4	-	-	-	21%/32%	13%/48%
Lab#5	-	-	-	-	23%/23%

Figure 4. Profile of 20 patient CSF samples analysed in six laboratories across Europe. (a) Average CV for each sample analysed in six laboratories plotted against sample concentration (µg/ml) shows that the samples with the highest NfL content had the lowest CV. **(b)** Differences between OD values of the standard curve do not explain the high inter-laboratory CV of the samples. The table shows mean per cent difference of OD values of the standard curve and the mean per cent difference of the calculated NfL concentrations.

Parameters	Lab #1	Lab #2	Lab #3	Lab #4	Lab #5	Lab #6
Sample storage prior to analysis (^{o}C)	-80	-80	-80, 4	-80	-80	-80
ELISA kit storage prior to analysis (^{o}C)	4	4	4	4	2-8	2-9
Temperature (^{o}C) at which ELISA was performed	21-22	21-22	22-24	20-24	21-22	20-24
Pipettes calibrated prior to experiment	No	Yes	Yes	No	No	No
100-200 µl pipette tips used for loading samples	Yes	Yes	Yes	Yes	No	Yes
Time from standard preparation to the start of incubation on the plate (minutes)	20-60	< 60	< 60	< 30	15	15
Different pipettes used for preparing Standards	Yes	Yes	Yes	No	No	Yes
Plate agitation at which incubations were performed (rpm)	800	800	800	800	800	800
Time of colour reaction (minutes)	15	20	15	15	15	15
Plate protected from light during colour Reaction	Yes	No	Yes	No	Yes	Yes
No. of freeze/thaw cycles of samples before shipment	2	2	2	2	2	2
No. of shipments before samples assayed	1	2	2	2	2	2
Sample condition on arrival	Frozen	Frozen	Frozen	Frozen	Frozen	Frozen
Curve fitting model used for data analysis	4PL	4PL	4PL	4PL	4PL	4PL

Table 4. Sample processing and assay handling parameters. (Adapted from Petzold et al., 2010)