City, University of London Institutional Repository

Parallelism in neurodegenerative biomarker tests: hidden errors and the risk of misconduct

A. Petzold, Joachim Pum & David P. Crabb

View supplementary material

Published online: 19 Nov 2025.

Submit your article to this journal

Article views: 209

View related articles

View Crossmark data

INVITED REVIEW

Parallelism in neurodegenerative biomarker tests: hidden errors and the risk of misconduct

A. Petzold[a], Joachim Pum[b] and David P. Crabb[c]

[a]University College London, Queen Square Institute of Neurology, London, United Kingdom; [b]LABanalytics GmbH, Jena, Germany; [c]City St. George's, University of London, London, United Kingdom

ABSTRACT

Biomarkers are critical tools in the diagnosis and monitoring of neurodegenerative diseases. Reliable quantification depends on assay validity, especially the demonstration of parallelism between diluted biological samples and the assay's standard curve. Inadequate parallelism can lead to biased concentration estimates, jeopardizing both clinical and research applications. Here, we systematically review the evidence of analytical parallelism in body fluid (serum, plasma, cerebrospinal fluid) biomarker assays for neurodegeneration and evaluate the extent, reproducibility, and reporting quality of partial parallelism. This systematic review was registered on PROSPERO (CRD42024568766) and conducted in accordance with PRISMA guidelines. We included studies published between December 2010 to July 2024 without language restrictions. Eligible studies included original research assessing biomarker concentrations in body fluids with data suitable for evaluating serial dilution and standard curve parallelism. The data extraction for interrogating parallelism included dilution steps, measured concentrations, and sample types. For each study, we generated parallelism plots in a uniform and comparable way. These graphs were used to come to a balanced decision on whether parallelism or partial parallelism was present. The risk of bias was assessed based on sample preparation, buffer consistency, and methodological transparency. Of 44 eligible studies, 19 provided sufficient data for generating 49 partial parallelism plots. Only 7 plots (14%) demonstrated clear partial parallelism. Partial parallelism was typically achieved over a narrow dilution range of about three doubling steps. Most assays deviated from parallelism, risking over- or underestimation of biomarker levels if determined at different dilution steps. A high risk of bias was identified in 9 studies using spiked or artificial samples, inconsistent dilution buffers, or incomplete reporting. Several studies assessed sample-to-sample parallelism rather than sample-to-standard, contrary to guidelines by regulatory authorities. In conclusion, partial parallelism was infrequently observed and inconsistently reported in most biomarker assays for neurodegeneration. Narrow dilution ranges and variable methodologies limit generalizability. Transparent reporting of dilution protocols and adherence to established analytical validation guidelines are needed. This systematic review has practical implications for clinical trial design, regulatory approval processes, and the reliability of biomarker-based diagnostics.

1. Introduction

Biomarker assays have become central to advancing research and clinical applications in neurodegeneration. For instance, the astrocytic biomarker glial fibrillary acidic protein (GFAP) has been approved by the U.S. Food and Drug Administration (FDA) as part of a panel to guide decisions on brain imaging after trauma [1]. Similarly, the FDA granted rapid approval for novel treatments in multiple sclerosis (MS) [2] and amyotrophic lateral sclerosis (ALS) [3] on the basis of

neurofilament (Nf) biomarkers [4]. The accurate quantification of such biomarkers relies on robust assay performance. To ensure this, the FDA and other regulatory authorities have established comprehensive frameworks for biomarker validation, covering all stages from sample collection and processing to analytical validation, clinical application, quality control, and accreditation. However, the breadth of this framework can also create ambiguity. For example, the terms "sensitivity" and "specificity" are used differently in analytical versus clinical contexts (Table 1). This distinction is

Table 1. Framework for laboratory test validation as provided by regulatory authorities. Abbreviations: U.S. Food and Drug Administration = FDA, conformit'e europ'eenne = CE, College of American Pathologists = CAP, Clinical and Laboratory Standards Institute = CLSI, International Organization for Standardization = ISO, analytical measurement range = AMR, limit of blank = LoB, limit of detection = LoD, limit of quantitation = LoQ. Table adapted from reference [9] with a focus on quantitative laboratory tests.

Laboratory test	ISO 15189/17025		CAP	
	FDA approved/CE certified	In-house & modified FDA-approved/ CE-certified	FDA approved/CE certified	In-house & modified FDA-approved/CE-certified
Precision & Bias	Verify CLSI EP15-A3	Establish CLSI EP15-A3	Verify CLSI EP15-A3	Establish CLSI EP15-A3
Method Comparison Experiment	Compare with current method, CLSI EP09-A3 Regression Analysis, Difference Plot	Compare with reference method CLSI EP09-A3 Regression Analysis, Difference Plot	Compare with current method CLSI EP09-A3 Regression Analysis, Difference Plot	Compare with reference method CLSI EP09-A3 Regression Analysis, Difference Plot
Analytical Sensitivity	— —	Establish CLSI EP17-A2 LoB, LoD, LoQ, Precision Profile Approach, Probit Analysis	Verify Documentation from manufacturer or literature, CLSI EP17-A2 Verification of LoB, LoD, LoQ	Establish CLSI EP17-A2 LoB, LoD, LoQ, Precision Profile Approach, Probit Analysis
Analytical Specificity	— —	Establish CLSI EP07-A Test for hemolysis, icterus and lipemia, potential cross-reactivities	Verify Documentation from manufacturer or literature	Establish CLSI EP07-A Test for hemolysis, icterus and lipemia, potential cross-reactivities
Diagnostic Sensitivity	— —	— —	— —	— —
Diagnostic Specificity	— —	— —	— —	— —
Linearity	Verify Evaluation of linearity, Calibration Verification, Verification of AMR	Establish Evaluation of linearity, Calibration Verification, Establish AMR	Verify Evaluation of linearity, Calibration Verification, Verification of AMR	Establish Evaluation of linearity, Calibration Verification, Establish AMR
Carryover	Verify Standard protocol or Short protocol	Establish Standard protocol	Verify Standard protocol or Short protocol	Establish Standard protocol
Measurement Uncertainty	Establish ISO 11352/NORDTEST Method	Establish ISO 11352/NORDTEST Method	Establish ISO 11352/NORDTEST Method	Establish ISO 11352/NORDTEST Method
Reference Range/Cutoff	Verify Documentation from manufacturer or literature	Establish CLSI EP28-A3c Use direct or indirect method for data collection	Verify CLSI EP28-A3c (Use direct or indirect method for data collection) OR Documentation from manufacturer or literature	Establish CLSI EP28-A3c (Use direct or indirect method for data collection)

illustrated by neurofilament light chain (NfL), which has a reported analytical specificity of 99.3% [5] due to minimal cross-reactivity with other Nf isoforms, yet demonstrates low clinical specificity because blood NfL levels rise across a wide range of neurological diseases [2,3] and in many conditions that compromise the integrity and function of neurons and their connections [4,6].

The framework for laboratory test evaluation presented in Table 1 requires additional clarification because certain terms are frequently misinterpreted in the literature [7]. A notable example concerns the distinction between "parallelism" and "dilution linearity." Guidance documents such as ISO/IEC 17025 and EURACHEM refer to parallelism when illustrating method validation, but the concept largely originates from applied laboratory practice. In practical use, parallelism and dilution linearity are sometimes conflated because both involve serial dilutions. The key

distinction lies in the sample type: dilution linearity experiments use samples spiked with analyte at concentrations designed to minimize matrix effects, whereas spiked samples are not permissible for assessing parallelism [7,8]. Although this difference may appear subtle, as will be argued in the present critical review, it has important implications for assay validation and interpretation.

Beyond these terminological distinctions, core analytical parameters such as precision, accuracy, linearity, and the evaluation of matrix effects remain central to assay validation [9]. Precision and accuracy ensure that results are both reproducible and correct, while linearity confirms that measured values remain proportional to analyte concentrations across the assay's dynamic range [9]. These parameters can be compromised by matrix effects, which introduce systematic bias [7,10]. This challenge is particularly pronounced when analyzing complex biological samples such as serum and

plasma [7,9,10], especially in mass spectrometry-based assays [11–13]. To mitigate such issues, harmonization across laboratories and analytical platforms has become a major focus in biomarker validation studies [14–17]. The goal is to ensure comparability of results and reliability of clinical decision-making [7]. As emphasized earlier, one of the hidden sources of error is the lack of parallelism [18,19]. For this reason, regulatory authorities require demonstration of parallelism between serially diluted samples and the assay calibration curve [20,21]. Demonstrating parallelism ensures that measured concentrations remain consistent and accurate across the relevant dilution range. Yet, the extent to which guidelines [20,21] for testing parallelism are systematically applied in biomarker validation studies remains unknown.

Therefore, we systematically reviewed the biomarker literature on neurodegeneration for evidence of parallelism testing, beginning with the first report of its absence in the neurofilament heavy chain (NfH) ELISA, which was attributed to protein aggregate formation [18]. Protein aggregation is a hallmark of many neurodegenerative diseases [22], meaning that the lack of parallelism observed in NfH assays has implications far beyond a single biomarker [18]. Building on this observation [18], we examined subsequent studies to determine whether other biomarkers are similarly affected. Deviations from parallelism [23] can lead to systematic over- or underestimation of biomarker concentrations, creating risks for misinterpretation that are particularly consequential in the context of clinical trials [2,3] and regulatory submissions [1,20,21]. Finally, our critical review underscores the limitations of current reporting practices in biomarker validation studies, which often include inconsistent dilution protocols, incomplete reporting of dilution ranges, and reliance on non-representative spiked or artificial samples.

With regard to terminology, we begin by clarifying the concept of parallelism, a term originally rooted in geometry. Although statistical approaches to testing parallelism are available, a detailed discussion of these methods is provided in the supplementary materials. In the main text, we instead focus on visual methods, as they are more accessible to readers without a strong mathematical background. This approach is supported by illustrative examples to ensure clarity. We then present the methodology of our systematic review, conducted in strict accordance with the PRISMA 2020 guidelines. The resulting findings are interpreted in the context of clinical laboratory science and extended to address broader methodological and regulatory issues. Finally, we highlight practical recommendations for laboratories, manufacturers, and regulators, with the aim of ensuring applicability beyond the research community.

2. Terminology and concept

The formal concept of parallelism can be traced to the Greek mathematician, geometer, and logician Euclid (Εὐκλείδης), who, around 300BCE, authored the seminal treatise Elements. In Book I, Definition 23, Euclid defined parallel lines as "straight lines which, being in the same plane and being produced indefinitely in both directions, do not meet one another in either direction." However, it is Euclid's fifth postulate, known as the parallel postulate, that has historically attracted the most scrutiny. Over the centuries, mathematicians sought to prove the fifth postulate using Euclid's other axioms. These efforts persisted for over two millennia and produced many flawed or incomplete proofs, reflecting a significant historical misconception [24]. It was not until the nineteenth century that mathematicians like Carl Friedrich Gauss (1779–1844), Lobachevsky (1829–1830), and Bolyai (1832) demonstrated that entirely consistent non-Euclidean geometries could be constructed by replacing the fifth postulate with alternative versions [24,25]. In a letter to Bolayi (17-DEC-1799) Gauss wrote: "It is true that I have come upon much which by most people would be held to constitute a proof: but in my eyes it proves as good as *nothing*."

Similarly, in the context of present, critical, systematic review, parallelism, though seemingly straightforward at first glance, demands careful scrutiny. Just as for Euclid's parallel postulate, analytical parallelism in quantitative assays must not be assumed but must be rigorously tested, substantiated, *and importantly lack of parallelism must be understood*.

2.1. Parallelism and standard curves

The accurate determination of analyte concentrations using a standard curve requires that the dilution series of the test sample exhibits parallelism with the standard curve [26]. This means the two curves must be similar functions, differing only by a scaling factor along the dose axis (in present systematic review, this is always the y-axis), so that interpolation yields valid and reproducible results. Without demonstrated parallelism, calculations derived from a standard curve may lead to inaccurate quantification [26]. In its simplest form *interpolation* is linear. Hence, the framework to verify/establish linearity for a laboratory test is in Table 1. Linear interpolation is a method of curve fitting using linear polynomials to construct new data points within the range of a discrete set of known data points

[27]. It is absolutely crucial to recognize that, for bio-marker concentrations from samples, calculations are only permitted for data points within the range of the points of the standard curve. Expanding from linear interpolation, quadratic, cubic, four-parameter logistic (4PL) [28,29], 4PL with logarithmic scaling of the dose (y-axis) [30], and five-parameter logistic (5PL) [29,31–33] standard curves have entered contemporary laboratory routine. Extrapolation is not allowed [9,27,30]. It is mandatory to demonstrate parallelism for the selected type of standard curve and real-world samples [20].

Despite its importance, the validation of parallelism presents a practical challenge for many laboratory workflows [9]. The concept, as reviewed here and long known in analytical chemistry, has only more recently gained sustained attention in the context of regulatory method validation for biomarker assays [10,21]. Variability in internal standards (ISV) has been identi-fied as a core factor contributing to measurement error, further underscoring the need for robust validation.

Regulatory agencies have begun to address the issue of parallelism. The FDA, through its 2022 M10 Bioanalytical Method Validation (BMV) guidance, and the European Medicines Agency (EMA) both highlight parallelism as a critical validation criterion [20,34]. According to the EMA, parallelism is defined as follows: "Parallelism demonstrates that the serially diluted incurred sample response curve is parallel to the cali-bration curve" [20].

2.1.1. Spiked samples in parallelism

The use of spiked samples in bioassay validation is an interesting strategy to assess analytical performance, particularly in relation to matrix effects and assay spec-ificity [10]. The use of spiked samples is frequently employed in the context of mass spectrometry-based methods, where matrix-induced variability is a major concern to the accuracy of quantification [11,35]. Current guidelines suggest that variations of less than 20% are generally acceptable when comparing spiked samples with calibrators or quality controls [10,36,37]. However, this approach typically focuses on perfor-mance at fixed concentration levels and does not explicitly evaluate assay behavior across wider dilution ranges or in the presence of other factors that apply to real-world samples [37].

Therefore, one of the key limitations of spiked sam-ples is the requirement for high-concentration material to enable serial dilution and minimize matrix effects [37]. Although spiking can be informative under con-trolled conditions, it introduces several methodological concerns. First, the reference material used for spiking may differ structurally or functionally from the endog-enous biomarker. This includes the use of truncated peptide sequences or proteins derived from non-human species [38,39]. Second, the stability of spiked material can be suboptimal, particularly over long-term storage, where protein aggregation may occur [39]. Aggregation, a known source of error in parallelism assessment [18], cannot be adequately simulated using spiked samples.

Moreover, spiked samples do not replicate the com-plexity of physiological matrices. As noted in prior reviews, their utility is limited to evaluating internal standard variability introduced by physiological differ-ences between control and patient matrices [37]. Even when the spiked analyte behaves additively with the endogenous biomarker, this assumption may not hold under all conditions, and further concerns regarding non-linear interactions persist [40].

For these reasons, real-world, patient-derived sam-ples remain the gold standard for assessing parallelism [7,8]. Their use better reflects the biochemical and bio-physical properties relevant to clinical measurement [41]. In the context of this systematic review, the use of spiked samples is discussed as a potential source of bias.

2.2. Determination of parallelism

Visual assessment offers a complementary and often more intuitive approach to evaluating parallelism in bioassays [26]. Unlike formal statistical testing (see supplementary materials), graphical methods allow analysts to inspect the behavior of dilution curves directly, which is particularly valuable in the presence of noisy or incomplete datasets. Statistical expertise required to properly interpret hypothesis testing meth-ods [42–45] is not commonly included in the core training of laboratory scientists, clinicians, or regula-tory reviewers. This gap may lead to an overreliance on statistical outcomes, potentially overlooking mean-ingful deviations in curve behavior across extended dilution ranges [23]. In real-world applications, com-plete parallelism is rare, while partial parallelism is often observed within limited dilution ranges [23]. This practical observation has led to the operational definition of *partial parallelism*, acknowledging that assays may behave acceptably within specific, pre-defined ranges without meeting idealized criteria across the entire curve [46].

The value of visual methods for detecting lack of parallelism has been recognized in the literature since the landmark paper by Plikaytis [26], and has been expanded by subsequent studies [23,47–49]. Figure 1 illustrates the most common patterns observed.

The interpretation of the partial parallelism plots goes back to Euclid's geometry. The two lines in the partial parallelism plot need to be parallel. To be more precise, they need to be parallel to each other, on the y-axis, for at least part of the graph (Pattern 3 in Figure 1). Hence, the term *partial parallelism*. A vertical offset is permitted, which is also entirely consistent with the requirements for statistical testing [43]. Interpretation of partial parallelism plots is best if data for the standard curves are present. For the purpose of this systematic review, partial parallelism plots will be created for each study included. Based on the graphical interpretation, which are presented for each study, a binary decision will be made: presence (see Figure 1 pattern 3) or absence (Figure 1 patterns 1&2, and other patterns of deviation from parallel alignment) of partial parallelism.

3. Clinical context

The clinical implications of reporting artificially elevated biomarker concentrations due to lack of parallelism (for example, Pattern 1 in Figure 1) can be illustrated using neurofilaments. Both NfL [50,51] and NfH [52,53] have been established as valuable prognostic biomarkers in patients with Guillain-Barré syndrome (GBS) [54,55].

GBS is characterized by evolving paralysis, which in severe cases can compromize respiratory function and necessitate intensive care unit (ICU) admission with mechanical ventilation. While most patients experience transient paralysis followed by substantial recovery, a subset develop long-term disabilities, including the permanent loss of ambulation. Early identification of high-risk patients would therefore be of great clinical value, enabling personalized management strategies ranging from ICU admission decisions to treatment escalation.

Targeting high-efficacy, but often more costly, treatments to patients at greatest risk offers the dual benefit of optimizing clinical outcomes and improving resource allocation. Enhanced functional recovery in this subgroup would not only improve patient quality of life but also reduce downstream healthcare costs, including those associated with long-term disability, loss of work capacity, increased care needs, and broader economic deadweight losses. The importance of such strategies is reflected in World Health Organization (WHO) guidelines developed in collaboration with the Universal Health Coverage Partnership (UHC Partnership). Their report, covering 115 countries and representing more than three billion people, explicitly recognizes the essential role of clinical laboratory services in healthcare delivery [56]. While the report does not specifically address technical parameters such as parallelism, this level of detail is beyond its scope; it implies that such issues may be
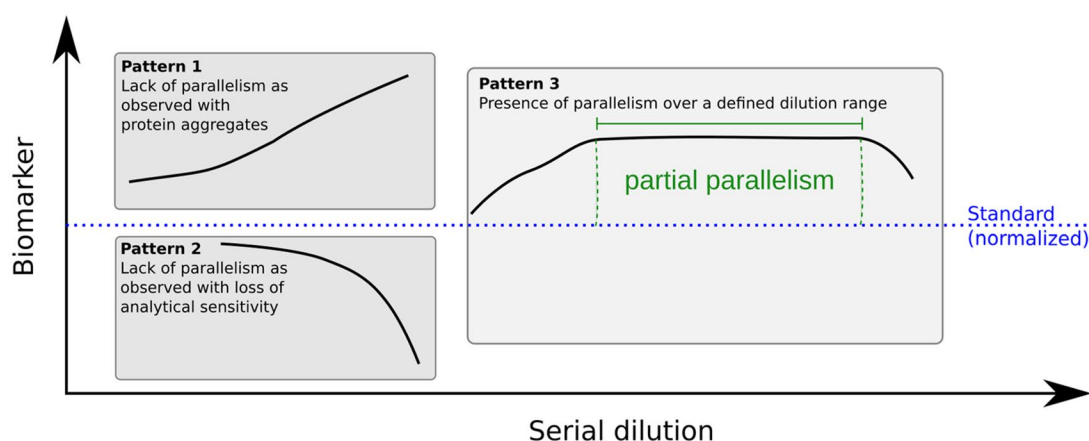

Figure 1. Visual assessment of parallelism. The figure illustrates characteristic patterns used to compare dilution ranges (x-axis) of the assay calibration standard (blue dotted line) with theoretical sample responses (black lines). Such visual patterns provide an accessible means of determining whether parallelism is maintained or lost and form the basis of our approach for evaluating parallelism across the assays included in this critical and systematic review. Pattern 1 shows an apparent increase in biomarker concentration with greater sample dilution, a phenomenon typically associated with the release of biomarker molecules from protein aggregates [18]. Pattern 2 shows a decrease in measured biomarker concentration with successive dilution steps, commonly observed when concentrations approach the assay's lower limit of detection (LoD; see Table 1). This underscores why regulatory authorities require validation of analytical sensitivity [9]. Pattern 3 demonstrates that parallelism may be achieved within a limited dilution range but lost again at higher dilutions. This pattern is termed partial parallelism [23]. As highlighted throughout this review, observed parallelism is most often partial.

incorporated under the remit of National Quality Control Laboratories [56].

To illustrate how the lack of demonstrated parallelism can have wide-ranging consequences, we highlight one recent example from the literature of NfL in GBS [57]. In this study, NfL concentrations were measured using a commercially available assay. The manufacturer recommends quantification at a dilution of 1:4. However, for a subset of samples, measurements were performed at a much higher dilution (1:400) without demonstrating parallelism between the diluted samples and the calibration standard across this extended range. The inferred implications are substantial:

- Inflated biomarker concentrations: Reported values ranged from approximately 1 pg/mL to nearly 10,000 pg/mL, which is several orders of magnitude higher than expected under either physiological variation or disease-related pathology.
- Potential patient misclassification: Elevated NfL levels were associated with more severe phenotypes defined by electrodiagnostic criteria [54,58]. The data imply that artificially inflated concentrations may have been interpreted as markers of irreversible axonal degeneration.
- Bias in predictive modeling: Predictive models are invaluable for risk stratification, but if their statistical significance arises from artificially elevated biomarker concentrations in a subgroup of severely affected patients, these models are unlikely to replicate in independent cohorts, clinical trials, or clinical practice.
- Impact on treatment evaluation: The study reported no treatment effect on NfL levels. One clinical interpretation might be to withhold a potentially effective therapy. However, the extreme variability in reported concentrations likely increased noise and obscured real effects, deviating from prior treatment trials that adhered to validated dilution ranges [2,3].
- Risk of deliberate misuse: A particularly concerning possibility arises in the context of treatment trials. If placebo samples were analyzed at a higher dilution range than treatment samples under Pattern 1 conditions (Figure 1), or at a lower range under Pattern 2 conditions, this would introduce systematic bias favoring a positive drug effect. Such practices would constitute scientific misconduct.

This example underscores, on multiple levels, the necessity of ensuring rigorous and reliable biomarker quantification. Inaccuracies not only risk patient misclassification but also undermine predictive modeling, treatment evaluation, and ultimately regulatory confidence. To evaluate how consistently this principle has been upheld in the field, we now proceed with a systematic review of the literature.

4. Methods

The protocol for this systematic review was submitted to the PROSPERO registry and published on the PROSPERO website. The study protocol can be searched under the registration number CRD42024568766. The 2020 PRISMA guidelines for reporting systematic reviews are followed [59]. The 27-item PRISMA checklist is uploaded.

4.1. Eligibility criteria

4.1.1. Inclusion criteria
All studies involving the analysis of parallelism of biomarkers in body fluids were considered for inclusion. No exclusion criteria were applied based on participant demographics or specific disease conditions. Body fluids included cerebrospinal fluid (CSF), urine, saliva, and other relevant bodily fluids. Various analytical techniques were considered, including immunoassays, mass spectrometry, ELISA, and other relevant methods for quantitative biomarker detection [11–13,21]. Data was permitted to be derived from purely analytical studies, experimental studies or clinical settings.

4.1.2. Exclusion criteria
Studies not involving biomarker sampling from body fluids. Research focused solely on tissue biomarkers or biomarkers derived from non-fluid sources. Articles lacking detailed methodology or insufficient data on biomarker sampling techniques [9]. Case reports, reviews, and opinion articles without original research data.

4.2. Information sources

Two databases were searched, Medline and Google Scholar.

4.3. Search strategy

A search of the MEDLINE database was conducted covering the period after publication of the lack of parallelism for NfH [18]. The dates entered to the search strategy were between 9-Dec-2010 and 12-July-2024. There were no language restrictions. The Entrez

Programming Utilities (E-utilities), provided by the National Center for Biotechnology Information (NCBI), were used in a Python script, which can be downloaded from the PROSPERO register.

4.3.1. Search strategy details

The search terms for the biomarkers (first search term) were: "neurofilament", "neurofilaments", "tau protein", "T-tau", "P-tau", "glial fibrillary acidic protein", "amyloid beta", "ubiquitin C-terminal hydrolase 1", "neurogranin", and "YKL-40". The search terms for the analytical methods (second search term) were: "method", "development", "linearity", "parallelism", and "doubling dilution". The first search term and the second search term were combined individually. These combinations were coded in python. The complete literature search code can be downloaded from the study protocol on the PROSPERO Registry under item number 17. The literature search was performed on 02-SEP-2024.

4.4. Study selection

Neurodegeneration is a prevalent feature in various human diseases, and biomarkers play a crucial role in its indirect assessment. Therefore, no specific disease-related restrictions were imposed. Limitations were focused on the analytical development of biomarker tests [9]. Studies were selected based on the availability of quantitative body fluid samples.

4.4.1. Selection process

All studies identified underwent a full-text review and review of supplementary data where available.

4.4.2. Tools used

A spreadsheet was kept using the PubMed Identifier (PMID) for each study. Studies included and excluded after review were clearly marked as such, including the reason for that decision.

4.5. Data collection process and data items

4.5.1. Data extraction

The data were extracted by hand from the values provided for samples at each dilution step into a spreadsheet. If data were only available in a figure, the corresponding author was contacted by email, containing the PROSPERO study number, asking to share the raw data.

4.5.2. Author contact

Two email requests for sharing the data items required were sent to all corresponding authors. The first email was sent on 12-SEP-2024. The second email was sent to the non-responders on 12-OCT-2024.

4.5.3. Data items

The three specific variables for which data were collected included: (1) dilution step, (2) measured biomarker concentration, and (3) sample type (standard, body fluid, artificial/spiked).

4.6. Risk of bias in individual studies

The risk of bias in individual studies was evaluated by carefully reviewing the methodologies used in sample selection and preparation. Specifically, we assessed whether the samples were native or spiked with protein standards and whether both the samples and standards were diluted using the same buffer, ensuring that it was indeed the correct dilution buffer [9]. Additionally, we recorded the selection of appropriate disease and control samples, noted whether the laboratory analyst was blinded during the experiments, and considered the reproducibility of the experiments through repeat assessments.

4.7. Summary measures

The summary measures used in this review include the deviation from the line of unity, which is represented by the horizontal line at a y-value of one in the partial parallelism plots [23]. The data are categorical, indicating the presence or absence of partial parallelism (yes/no). For studies where parallelism is observed, continuous data were be presented, specifically the dilution range within which partial parallelism was demonstrated. These two effect measures will be analyzed for each sample individually and, when meaningful ($n > 5$), also as a group mean.

4.8. Synthesis of results

The synthesis of results is primarily visual, utilizing partial parallelism plots [23] derived from the included studies. These visual representations are then summarized in table format and complemented by a narrative synthesis.

4.9. Risk of bias across studies

The primary focus is on determining whether partial parallelism exists between a sample from an individual with a specific disease and the protein standard used in the test. Practically, this requires providing data for a dilution series of the sample into the assay buffer (the same dilution

buffer that is used for the standard curve) [9]. Therefore, a high risk of bias was assigned if no such sample was available and instead, an artificial sample was created by spiking the body fluid collected with the protein. Similarly, if different buffers were used for dilution between the sample and the standard, the bias was also rated as high. A moderate risk of bias was assigned when an appropriate disease sample was not used, or if the laboratory analyst was not blinded, or if data on reproducibility were missing. In all other cases, the risk of bias was rated as low.

5. Results

5.1. Study selection

The flow diagram in Figure 2 outlines the process of study selection, detailing the number of records identified, included, and excluded [60]. A comprehensive literature search initially yielded 39 articles for further evaluation [16,35,61–97].

Additionally, a search on Google Scholar, using the same terms, identified five more references. [5,98–101] 5,98–101 All 44 articles,[16,35,61–99], including supplementary analyses where available, were thoroughly reviewed. The corresponding author was contacted twice by email to share data if this was necessary for creating partial parallelism plots. For four articles, no presently valid author contact details could be found [69,70, 73,75]. For the remaining papers, a careful examination of all data available resulted in the exclusion of 25 articles [78,79,80,81,82,83,84,85,86,87,88,70,89,90,91,92,93,94,73,95,75,76,96,16,97] due to the absence of data necessary for generating partial parallelism plots as per our predefined protocol.
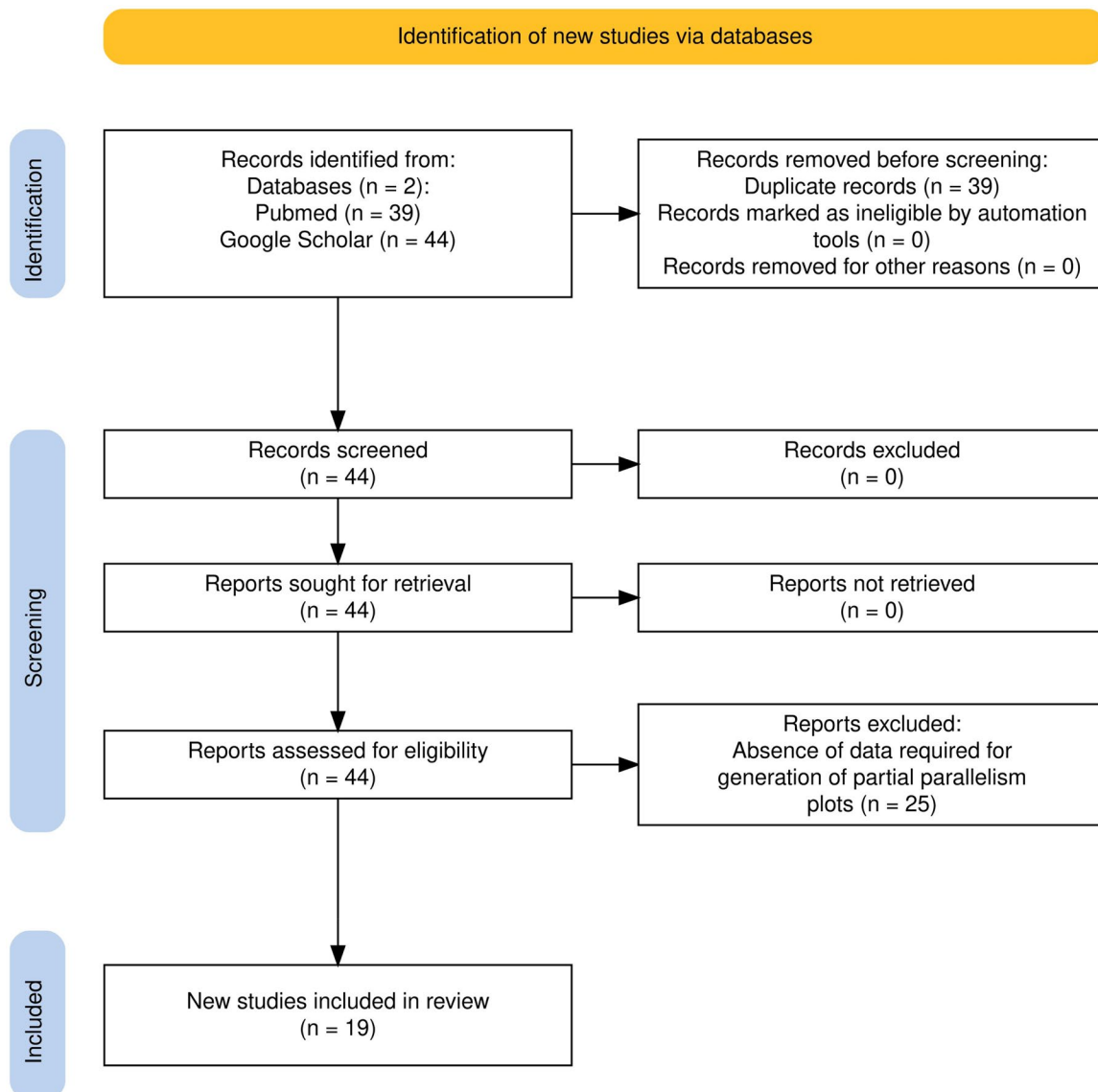
Figure 2. PRISMA flow diagram for the literature search, based on the PRISMA template code [60].

Consequently, 19 studies were selected for further evaluation [5,35,61–69,71,72,74,77,98–101].

5.2. Study characteristics

The characteristics of the included studies are summarized in Table 2. The majority of studies focused on human samples, primarily using plasma or CSF, with only a few investigating serum. Four studies examined brain tissue homogenates, while two used rodent tissue samples. All studies included control samples, and most also incorporated samples from a variety of disease conditions, with Alzheimer's disease (AD) being the most frequently studied. The biomarkers analyzed across the studies included the Amyloid β fragments, $A\beta_{1-40}$, $A\beta_{1-42}$, $A\beta$ oligomers, α-synuclein, DJ-1, total tau protein (Tau), phospho tau proteins (pTau181, pTau217, pTau231), Apolipoprotein E (ApoE) isoforms E2, E3, E4, neurofilament light chain (NfL), neurofilament heavy chain (NfH), GFAP, and poly(GP).

5.3. Risk of bias within studies

The risk of bias across the included studies is summarized in Table 3. Only one study was assessed as having a low risk of bias [61]. This study provided clear documentation on key factors, including blinding and sample handling.

Eight studies were rated as having a moderate risk of bias [5,35,63,64,66,74,98,100], mainly due to incomplete information regarding blinding procedures, particularly in relation to the blinding of the analyst.

Another eight studies were categorized as having a high risk of bias [62,65,67,68,71,72,77,99]. This rating was due either to the use of spiked samples [62,65,67,68,72,77,99], or due to incomplete documentation on whether the samples were spiked and how they were processed [71].

5.4. Results of individual studies

Detailed results for each included study are summarized in Table 4.

5.4.1. Amyloid β

The amyloid cascade hypothesis, proposed by Hardy and Higgins [102], has positioned amyloid β ($A\beta$) as a central pathological driver in AD, following proteolytic cleavage of the amyloid precursor protein (APP). Quantification of $A\beta$ peptides, particularly $A\beta_{1-42}$ and $A\beta_{1-40}$, and their ratios has since become integral to

Table 2. Study characteristics of the included studies, 61, 68, 100, 101, 35, 63, 64, 66, 5, 74, 98, 62, 65, 77, 69, 71, 72, 99, 67 sorted for biomarkers tested. Column abbreviations: P=plamsa, S=serum, C=CSF, T=tissue, H=human, R=rodent. Remaining abbreviations in alphabetical order: Aβ=amyloid beta, AD=Alzheimer disease, ALS=Amyotrophic Lateral Sclerosis, ApoE=apolipoprotein E, ctrl=control, ELISA=enzyme linked immonosorbent assay, elecsys=ElectroChemiLuminescence (ECL) technology for immunoassay analysis (Roche), FTL/FTLD=Fronto Temporal Lobar Dementia, GFAP=glial fibrillary acidic protein, IMR=immunomagnetic reduction, LC-MS/MS=liquid chromatography-mass spectrometry/mass spectrometry, MALDI-TOF=matrixassisted laser desorption/ionisation time of flight, MCI=minimal cognitive impairement, MS=multiple sclerosis, MSD=meso scale discovery, NfH=neurofilament heavy chain, NfL=neurofilament light chain, PD=Parkinson disease, ptau=phospho-tau, simoa=single molecule arrays, sAPP=soluble amyloid-β precursor protein, sFIDA=surface-based fluorescence intensity distribution analysis, VD=vascular dementia. Studies are sorted alphabetically according to the biomarker analyzed.

Ref	P	S	C	T	H	R	Disease(s)	Assay	Biomarker
98	x	–	–	–	x	–	Ctrl	HISCL	$A\beta 1-40$, $A\beta 1-42$
65	x	–	–	–	x	–	Ctrl	LC-MS/MS	$A\beta 40$, $A\beta_{1-42}$
69	x	–	x	–	x	–	Ctrl, AD	ELISA	$A\beta 1-40$, $A\beta 1-42$
72	x	–	–	–	x	–	Ctrl, AD	Simoa & ELISA	$A\beta 1-42$
35	–	–	x	–	x	–	Ctrl, AD	ELISA	$A\beta 1-42$
64	x	–	–	–	x	–	Ctrl	MALDI-TOF MS	$A\beta$ peptides
62	x	–	x	–	x	–	Ctrl	sFIDA	$A\beta$ aggregates
67	x	–	–	–	x	–	Ctrl, VD, PD, MCI, FTL, AD	sFIDA	$A\beta$ oligomers
68	–	–	x	–	x	–	Ctrl	MSD multi-array	α-synuclein, $A\beta_{1-42}$, DJ-1, total tau
63	–	–	x	–	x	–	ApoE genotypes	LC-MS/MS	ApoE (E2, E3, E4)
61	x	x	–	x	–	x	Ctrl, TBI, AD	MSD	GFAP
5	x	x	x	–	x	–	Ctrl, MS, ALS	ADVIA&Atellica	NfL
66	–	–	x	x	x	–	Ctrl	ELISA, Luminex	NfL, NfH
100	–	–	x	–	x	–	Ctrl, C9or7f2	Simoa	poly(GP)
77	x	–	x	x	x	x	Ctrl, AD	MSD, IMR	tau
71	–	–	x	–	x	–	Ctrl, AD	Elecsys	tau, Tau-p181
99	x	–	x	–	x	–	Ctrl, MCI, AD, FTLD	Simoa	Tau-p181-p231
74	–	–	x	x	x	–	Ctrl, AD	Simoa	Tau-p217
101	x	–	–	–	x	–	Ctrl, AD	Simoa	tau-p217, tau-p231

Table 3. Risk of bias assessment of the studies included. Studies are sorted according to the level of bias risk.

Ref	Bias risk	Explanation for risk of bias assessment
[61]	Low	Native & spiked samples used & investigator blinded
[101]	Moderate	Native samples used, but not stated if the two analysts blinded
[68]	Moderate	Native & spiked CSF samples used, but not stated if investigator blinded
[35]	Moderate	Native & spiked samples used, but not stated if analyst blinded
[63]	Moderate	Native samples used, but not stated if the analyst blinded
[64]	Moderate	Native samples used, but not stated if the analyst blinded
[66]	Moderate	Native & spiked samples used, but not stated if analyst blinded
[5]	Moderate	Native & spiked samples used, but not stated if analyst blinded
[74]	Moderate	Native & spiked samples used, but not stated if analyst blinded
[98]	Moderate	Native samples used, but not stated if the analyst blinded
[62]	High	Spiked samples used
[65]	High	Spiked samples used for high concentrations
[77]	High	Spiked samples of health donor plasma purchased from vendors, unknown how they were processed. Parallelism was done with spiked buffer only.
[69]	High	Spiked samples used
[71]	High	Leftover samples purchased from vendors, unknown whether or not they had been spiked or how they were processed.
[72]	High	Pooled plasma samples first immunodepleted & then artificially spiked with amyloid-beta
[100]	High	Native & spiked samples used & analyst blinded, but dilution not done in buffer
[99]	High	Spiked samples used
[67]	High	Spiked samples of human plasma & spiked PBS solution used

Table 4. Partial parallelism plots (PP-plots) from studies with available raw data, obtained either through publication or direct email request (details provided in the methods section). Comprehensive PP-plots for all included studies are presented in Figures 3–21. The specific dilution ranges demonstrating partial parallelism are explicitly indicated. Studies are sorted as in Table 3.

Ref	PP-plot	Partial parallelism achieved	Dilution range
[61]	Figure 3	Yes	1:4–1:16
[68]	Figure 4	No	—
[101]	Figure 5	Yes (Lilly pTau217 assay)	1:8–1:32
		Not for the pTau181 assays	—
		Not for the pTau231 assay	—
[35]	Figure 6	No	—
[63]	Figure 7	No	—
[64]	Figure 8	No	—
[66]	Figure 9	No	—
[5]	Figure 10	Yes	1:2–1:8
[74]	Figure 11	No	—
[98]	Figure 12	No	—
[62]	Figure 13	No	—
[65]	Figure 14	No	—
[77]	Figure 15	No	—
[69]	Figure 16	No	—
[71]	Figure 17	Yes (pTau181)	1:1.43–1:2.5
		Not for the total Tau assay	—
[72]	Figure 18	Yes	1:4–1:16
[100]	Figure 19	Yes	1:4–1:16
[99]	Figure 20	No	—
[67]	Figure 21	No (EDTA)	—
	Figure 21	Yes (Citrate, Heparin)	1:100–1:10,000

subsequent revisions of diagnostic criteria for AD [103–105]. However, aggregation-prone properties of Aβ [106,107] complicate immunoassay quantification, as epitope masking and altered conformations impair antibody recognition and disrupt dilutional parallelism [18]. On review of individual partial parallelism plots, this emerges as a consistent and persistent analytical problem:

- **Aβ_{1-40}**: None of the assessed immunoassays achieved partial parallelism across three independent studies [65,69,98]. For each study, deviations from expected parallelism were shown in the partial parallelism plots (see Figures 12, 14, and 16).
- **Aβ_{1-42}**: Likewise, most studies failed to demonstrate partial parallelism [35,65,68,69,98]. A single study reported successful parallelism [72]; however, this study was flagged as high risk for bias (Table 3), as it involved spiking immunodepleted pooled plasma with synthetic Aβ, a method that may not replicate the conformational complexity of endogenous peptides (see Figures 6, 4, 12, 14, 16, and 18).

- **Other Aβ species**: No evidence of partial parallelism was found for truncated Aβ peptides [64], aggregated forms [62], or oligomeric Aβ in EDTA plasma samples [67]. However, in the latter study, the use of citrate or heparin as anticoagulants enabled partial parallelism, highlighting the importance of pre-analytical variables for future immunoassay development (see Figure 21).

Although current diagnostic strategies emphasize Aβ cleavage products, APP itself remains a biomarker of interest [108]. Recent attention has turned to its soluble fragments (sAPPα, sAPPβ) [109], though the only identified assay for these analytes [75] could not be included in this systematic review. Together, these findings underscore ongoing analytical challenges in Aβ biomarker development. Despite three decades of intensive research [102], robust quantification of circulating Aβ, especially Aβ_{1-42}, remains elusive.

5.4.2. α-Synuclein

α-Synuclein remains a central candidate biomarker in the differential diagnosis of movement disorders, particularly PD and other synucleinopathies [110–113]. However, its intrinsic biochemical properties pose
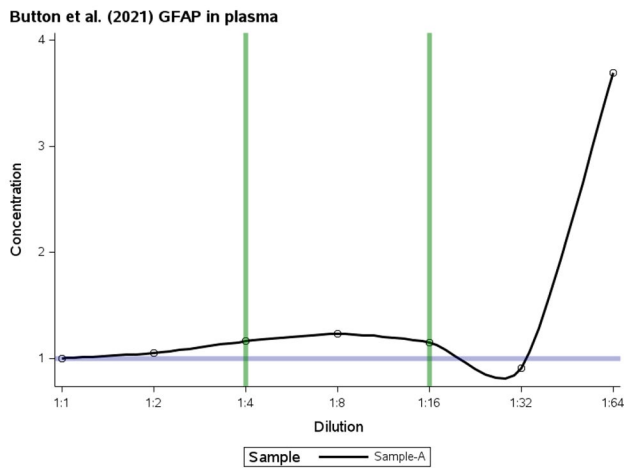
Figure 3. Partial Parallelism Plot for endogenous GFAP from plasma samples. The partial parallelism plot illustrates the average GFAP concentrations from three native plasma samples, based on the data summarized in table 7 of reference [61]. The individual sample data are shown in Figure 4D of the same reference [61]. Partial parallelism is evident for dilution steps ranging from 1:4 to 1:16. However, at the final dilution step of 1:64, while GFAP remains detectable by the assay, the partial parallelism plot clearly indicates a loss of parallelism.

substantial challenges for quantitative assay development. Notably, α-synuclein exhibits a pronounced tendency to aggregate [110,114], a characteristic that directly impairs immunoassay performance by compromising parallelism [18]. In line with these known biophysical limitations, the present review identified a failure to achieve partial parallelism in the only study that evaluated this criterion using standard dilution series [68] (see Figure 4).

5.4.3. DJ-1

DJ-1 (PARK7) has been implicated in the pathophysiology of both familial and sporadic Parkinson's disease, with several studies reporting elevated levels in patient cohorts [111,115]. Despite its biomarker potential, the analytical validation of DJ-1 assays remains insufficient. Notably, published work utilizing in-house Luminex-based immunoassays did not report formal testing for dilutional parallelism [111]. Within the current review, the only available dataset assessing this parameter demonstrated a lack of partial parallelism [68] (see Figure 4).
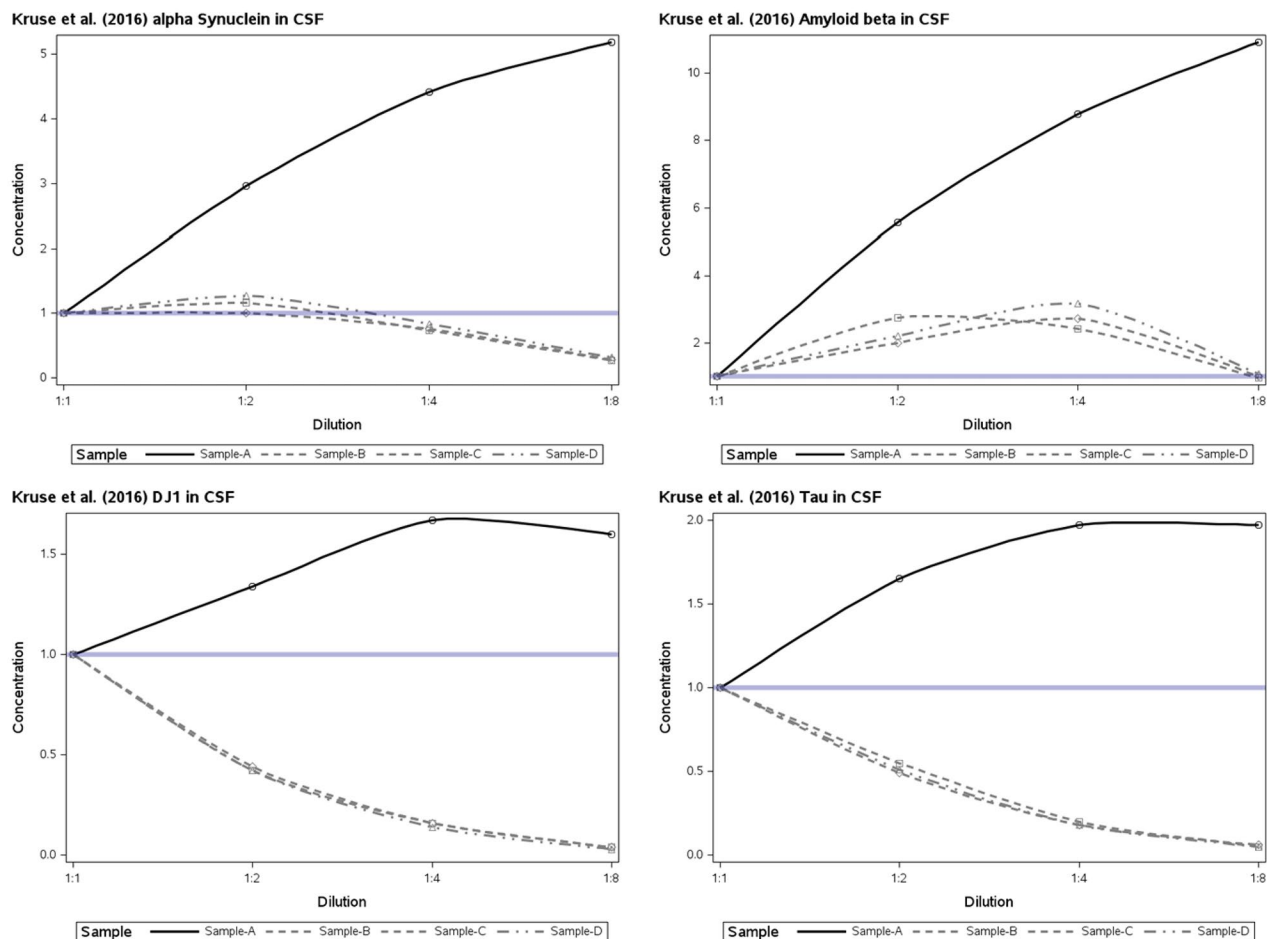
Figure 4. Partial parallelism plots for four proteins: α-synuclein, $A\beta_{1-42}$, DJ-1, total tau. Raw data were obtained from Supplementary Table S5 in reference [68]. The plots demonstrate a lack of partial parallelism across the tested dilution ranges.
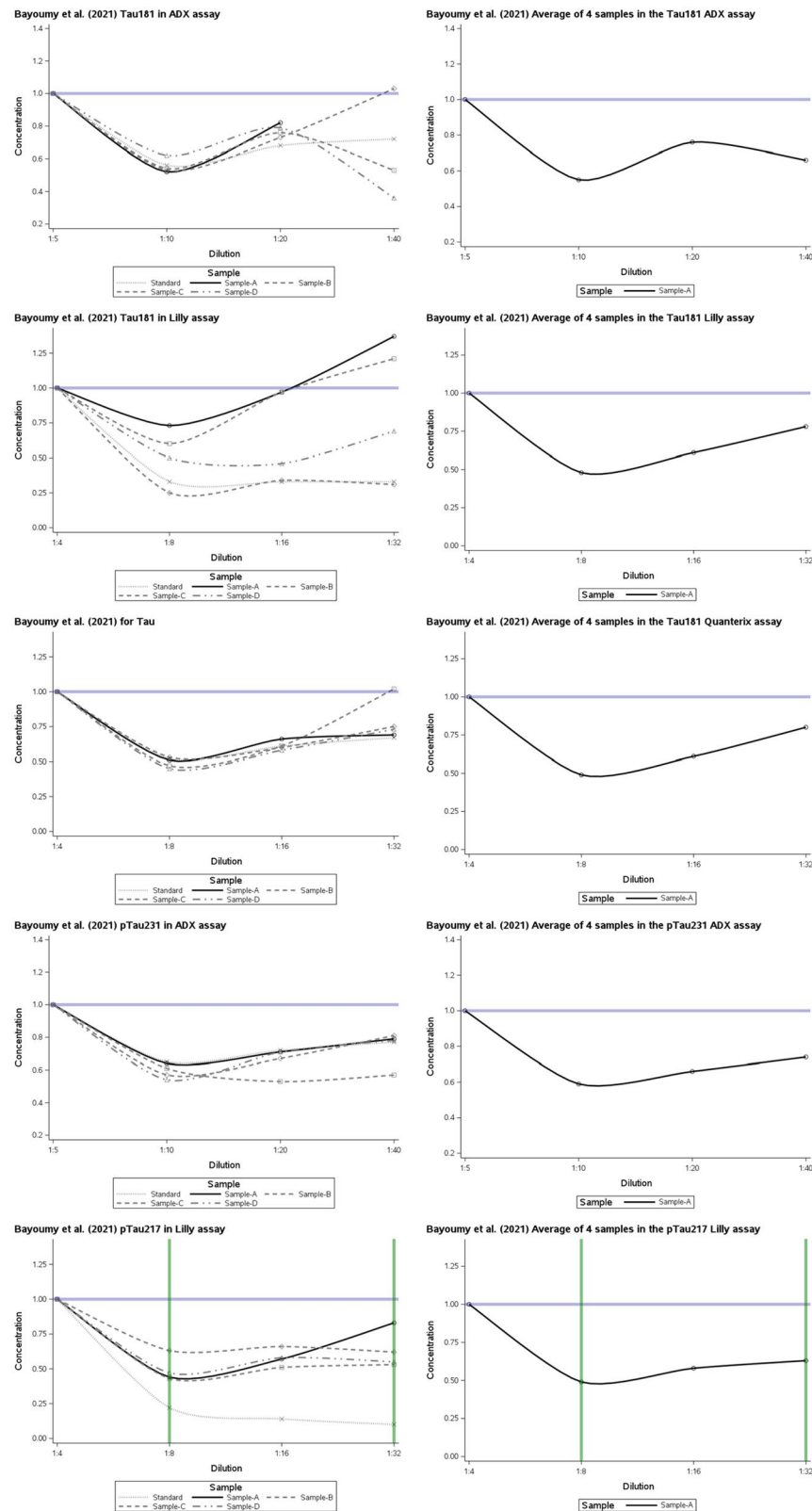
Figure 5. Partial parallelism plot for tau from native plasma samples [101]. Raw data were kindly provided by the authors. The plots show a lack of partial parallelism for all pTau181 and pTau231 assays across the tested dilution ranges. In contrast, partial parallelism is observed for the Lilly pTau217 assay between dilutions of 1:8 to 1:32.

5.4.4. Tau protein

The incorporation of CSF tau into the diagnostic criteria for AD represented a critical milestone in the development of neurodegenerative biomarkers [103]. Early diagnostic efforts predominantly focused on total tau, supported by robust clinical data [116]. However,
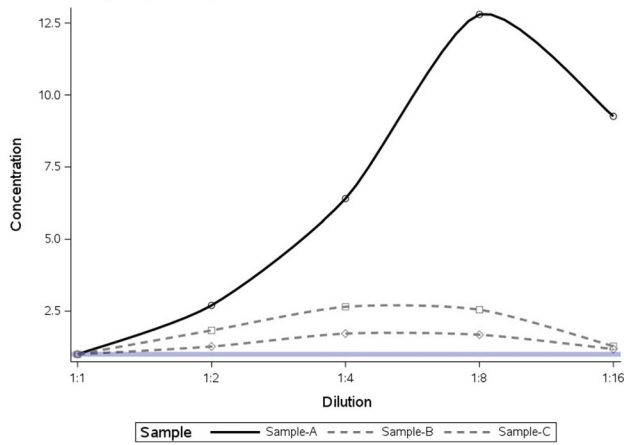
Figure 6. Partial Parallelism plot for $A\beta_{1-42}$ in CSF. Data were taken from Figure 5 of reference [35]. Datapoints were measured in pixel distances from that Figure. The plots demonstrate a lack of partial parallelism across the tested dilution ranges.

inter-laboratory variability in cutoff values ranged from 4 fmol/mL to 1140 pg/mL [117–121]. This highlighted challenges for assay validation [122,103].

Subsequent advances in antibody technology targeting specific phospho-epitopes of tau have facilitated the development of more analytically rigorous immunoassays. Among these, assays for phosphorylated tau at threonine-181 (p181), threonine-217 (p217), and threonine-231 (p231) are now widely utilized [123]. Notably, plasma p217tau currently demonstrates the highest diagnostic performance, with sensitivity and specificity comparable to CSF-based tau tests [103,116,123,124].

The results for parallelism of these assays are summarized as follows:

- **Total Tau:** Partial parallelism was not achieved in any of the reviewed studies [68,71,77], indicating persistent limitations in dilutional
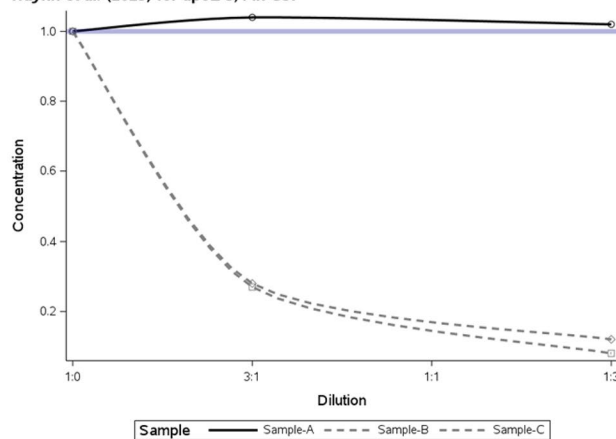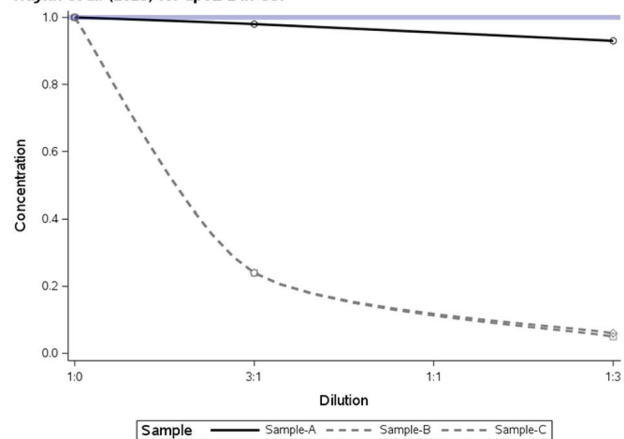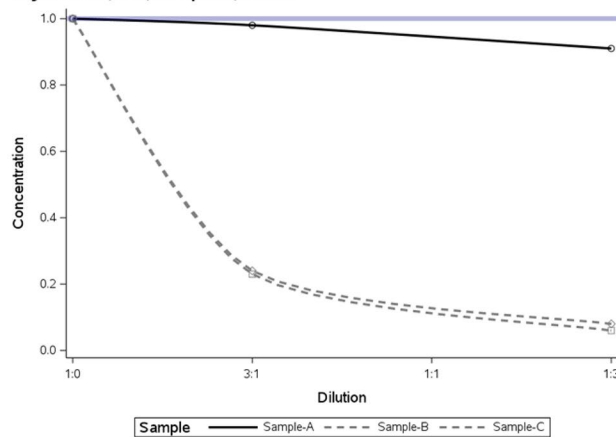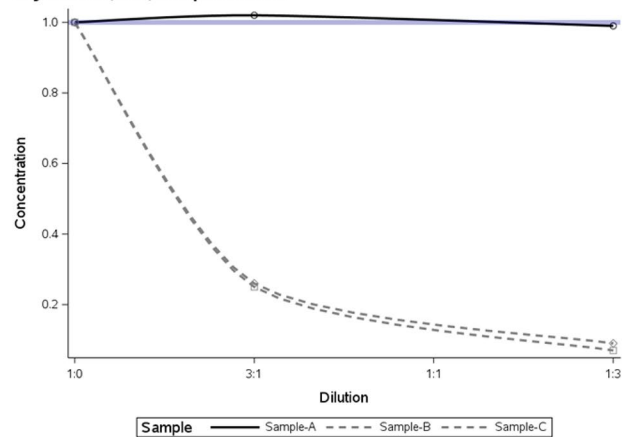
Figure 7. Partial Parallelism Plot for ApoE isoforms in CSF samples. This plot depicts the partial parallelism of apoE isoforms diluted in lumbar CSF, ventricular CSF, and bovine CSF. The data are derived from Figure 4 of reference [63]. The measured peptides include LAVYQAGAR (ϵ3&ϵ4, SNP Arg158), CLAVYQAGAR (ϵ2, SNP Cys158), LGADMEDVCGR (ϵ2&ϵ3, SNP Cys112) and LGADMEDVR (ϵ4, Arg112).
The plots demonstrate a lack of partial parallelism across the tested dilution ranges.
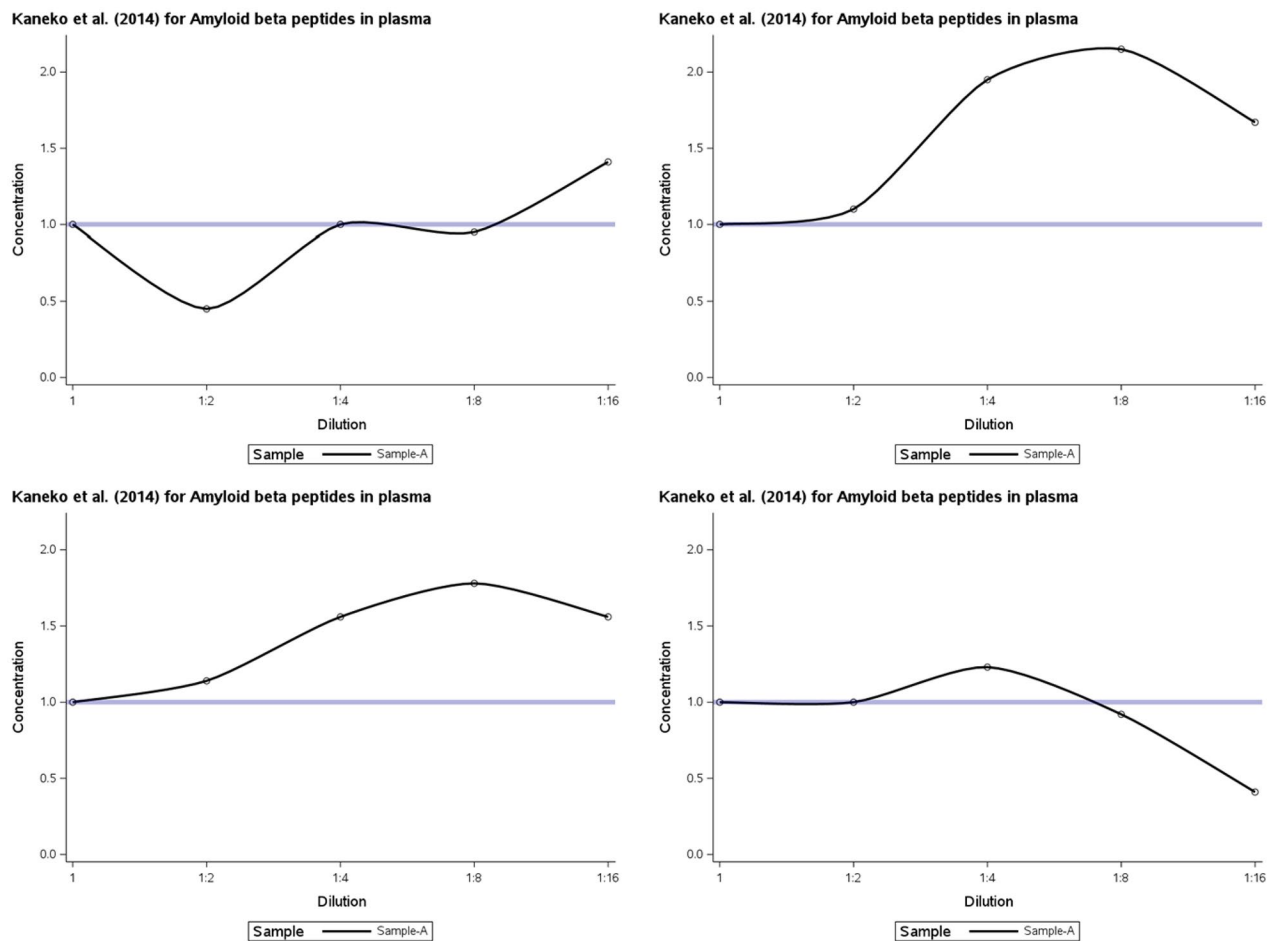
Figure 8. Partial Parallelism Plot for serial dilution of Amyloidβ peptides in plasma. This plot illustrates the partial parallelism for serial dilutions of various amyloid-β peptides in plasma. Data were analyzed across complete dilution ranges (five data points) for comparison in the partial parallelism plots. Such data were available for 4 out of 18 (22%) plots presented in Figure 4 of reference [64]. The peptides included Aβ_{6-40}, Aβ_{5-40}, Aβ_{1-38}, Aβ_{1-40}. The plots demonstrate a lack of partial parallelism across the tested dilution ranges.

linearity across all tests available. This lack of parallelism raises concerns about the quantitative reliability of total tau assays for individualized or longitudinal use (see Figures 4,15,17).

- **Phospho-Tau p181:** Partial parallelism was demonstrated in one study [71], but not in two others [99,101], suggesting inter-assay variability or dependency on specific analytical conditions (see Figures 5 and 20).

- **Phospho-Tau p217:** Achieved partial parallelism was reported in one study [101]; however, conflicting results from another study [74] may reflect noise or technical artifacts. The overall evidence remains promising but requires further investigation (see Figures 5 and 11).

- **Phospho-Tau p231:** No study to date has demonstrated satisfactory partial parallelism for p231-tau assays [99,101], pointing to persistent

limitations in current assay performance for this isoform (see Figures 5 and 20).

Taken together, these analytical findings support the emerging consensus that phospho-tau p217 is currently the analytically most robust and clinically best validated tau biomarker [101,123,124].

5.4.5. Apolipoprotein E

Apolipoprotein E (ApoE) alleles (E2, E3, E4) are well-established genetic risk factors associated with various neurodegenerative diseases [63,125–127]. In the assay evaluated [63], partial parallelism could not be demonstrated for any of the ApoE isoforms (see Figure 7). However, it is important to note that the analysis was based on only three samples, limiting the robustness of this conclusion. Notably, one sample (Sample A) did exhibit some degree of partial parallelism for the E3/4 and E4 variants. These preliminary

findings suggest potential for assay refinement, and further developmental work [63] with a larger sample set will be necessary to accurately assess dilutional behavior and parallelism across ApoE isoforms.

5.4.6. Neurofilament proteins

Among the five known neurofilament subunits [4], neurofilament light (NfL) and heavy (NfH) chains have been assessed [5,66]. For NfH, a five-parameter logistic
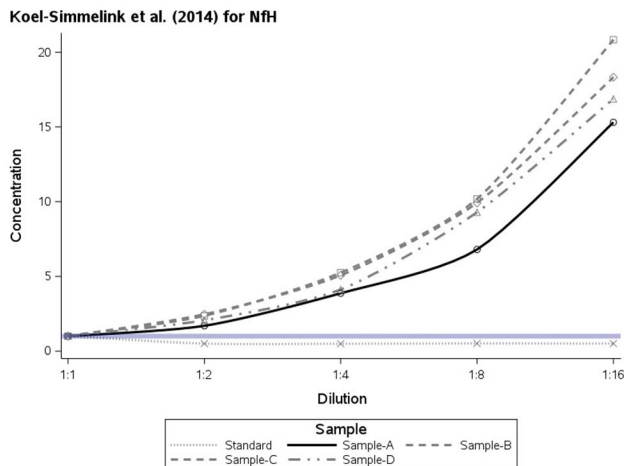
Figure 9. Partial parallelism plot for NfH quantification across different sample types. Sample A represents buffer spiked with NfH, while Samples B–D are derived from native CSF [66]. The raw data were provided by the corresponding author. The plot shows increasing NfH concentrations with increasing dilution steps, while the standard curve demonstrates partial parallelism only within the 1:1 to 1:16 dilution range. Overall, the results indicate a lack of partial parallelism across the full dilution range tested.

(5PL) model was employed to generate the standard curve [66]. However, the authors' conclusion that the assay demonstrated no evidence of non-parallelism is not fully supported by the partial parallelism plot based on four samples (see Figure 9).

In contrast, the NfL assay showed group-level evidence of partial parallelism across a dilution range of 1:2 to 1:8, as demonstrated in a study using ten samples [5] (see Figure 10). Despite these findings at the group level, deviations from partial parallelism remain evident at the level of individual patient samples. These inconsistencies remain to be addressed to enable reliable use of neurofilament measurements for personalized clinical decision-making [128].

5.4.7. Glial fibrillary acidic protein

In the GFAP assay reported by Butcher et al. [61] a minor discrepancy was observed between the claimed dilution range for parallelism (1:1 to 1:32) and the actual range showing partial parallelism (1:4 to 1:16; see Figure 3). The authors also noted a deviation from the recommended protocol [7] (see Table in reference 61), having used only 3 instead of 5 plasma samples for parallelism assessment. Notably, all three samples were derived from patients with acute traumatic brain injury (TBI) and exhibited high GFAP concentrations, potentially introducing a biomarker selection bias [129]. Previous studies have reported a lack of dilutional parallelism for GFAP in CSF in chronic conditions such as ALS [1,130,131]. Additionally, the pronounced deviation from the line of unity at a dilution of 1:64 observed in Figure 3 remains unexplained. However, given that earlier findings were based on different

Figure 10. Partial Parallelism Plot for Neurofilament Light Chain (NfL) quantification. The plot illustrates the quantification of NfL in serum samples spiked with 500 ng/mL of the peptide standard [5]. The graph to the left displays the individual dilution curves, while the graph to the right shows the averaged data from 10 spiked serum samples. The raw data were provided by the corresponding author. The graph demonstrates that parallelism is demonstrated, on a group level, for a dilution range of 1:2–1:8 for recombinant NfL in serum using the assay buffer.
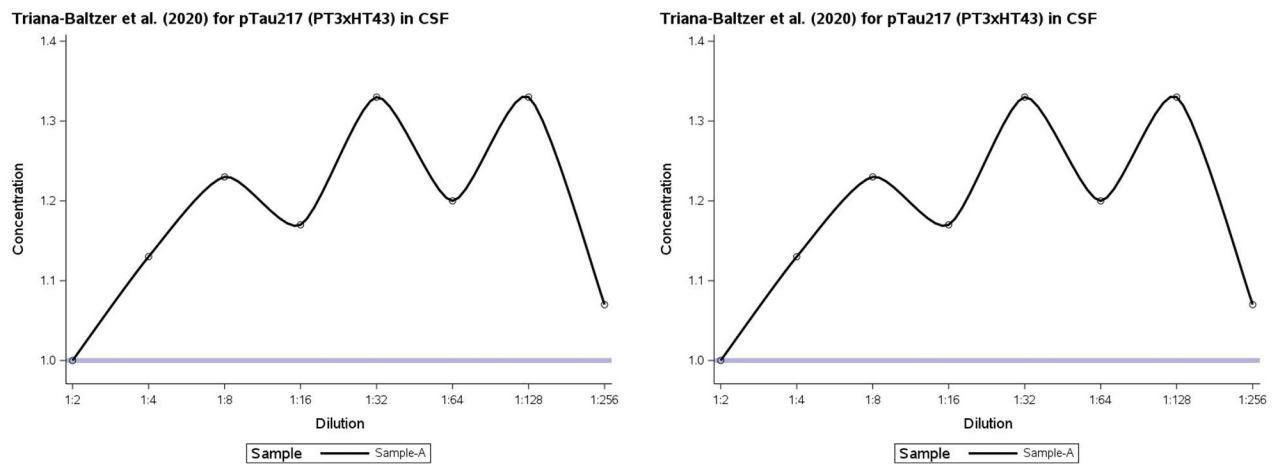
Figure 11. Partial parallelism plots for serial dilution of CSF measuring pTau217 using PT3xHT43 and PT3xPT82 assays. Data were extracted from Figure 2A in reference [74]. The plots do not demonstrate partial parallelism; however, this may be influenced by data noise. Averaged data from additional samples tested within the dilution range of 1:8 to 1:64 could provide further clarity.
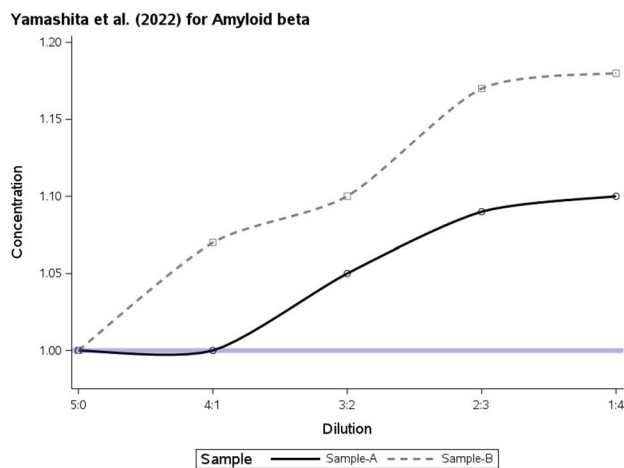

Figure 12. Partial Parallelism plot for $A\beta_{1-42}$ and $A\beta_{1-40}$ quantified from plasma samples spiked with the respective Amyloidβ peptides. The data were taken from supplementary Table S1 in reference [98]. The plot indicates a lack of partial parallelism across the dilution range (native sample to 1:4) tested.

ELISA platforms and assessed GFAP in CSF rather than plasma, further investigation is needed to determine whether partial parallelism can be reliably demonstrated across neurodegenerative diseases in addition to TBI.

5.4.8. Dipeptide repeat proteins

Among the five known dipeptide repeat proteins [poly(GP), poly(GA), poly(GR), poly(PR), and poly(PA)], poly(GP) has been quantitatively assessed [100]. The assay utilized a combination of custom-made rabbit polyclonal antibodies and a mouse monoclonal antibody (TALS 828.179). Due to heteroscedasticity observed in the signal across the standard curve, the

data required post-processing prior to curve fitting. Both 4PL and 5PL models were initially evaluated, with the 4PL model ultimately selected and weighted using 1/Y [2]. The test demonstrated a diagnostic sensitivity and specificity of 100% for distinguishing 40 individuals with *C9orf72* repeat expansions from 15 healthy controls [100]. Notably, partial parallelism was observed between dilutions of 1:4 and 1:16, a narrower dynamic range than the less diluted 1:1 to 1:2 concentration proposed by the authors as anchors for assessing dilutional parallelism (see Figure 19).

5.5. Synthesis of results

Partial parallelism plots were generated from 19 studies [61,68,100,101,35,63,64,66,5,74,98,62,65,77,69,71,72,99,67]. Several studies had analytical results necessitating the generation of more than one partial parallelism plot. In total, 49 partial parallelism plots were generated (Figures 3–21). Nine of these plots were on average data from experiments with more than five samples. The Figures illustrate that partial parallelism was achieved in only 7 out of 49 experiments (14%) (Table 4).

The dilution ranges demonstrating partial parallelism were typically narrow [5,61,71,72,100,101]. Narrow dilution ranges of about three dilution steps may be acceptable in laboratory assays investigating parallelism, where a narrow range can be expected [44]. But for the biomarkers reviewed here, the literature reports range over two to three magnitudes of concentration [132–137].

Notably, one study diluted serum pools from healthy controls (low NfL concentrations) into samples from individuals with multiple sclerosis (high NfL concentrations) instead of using assay buffer
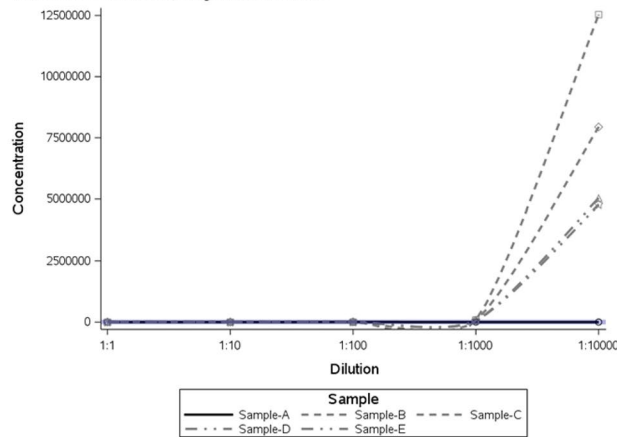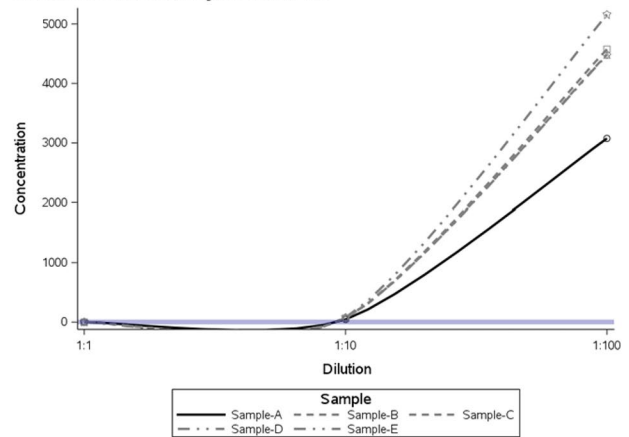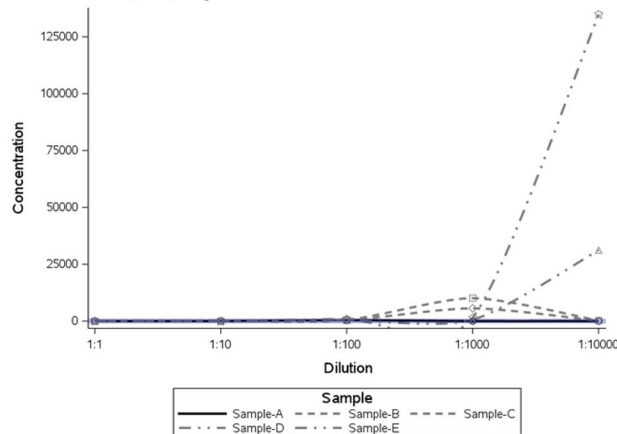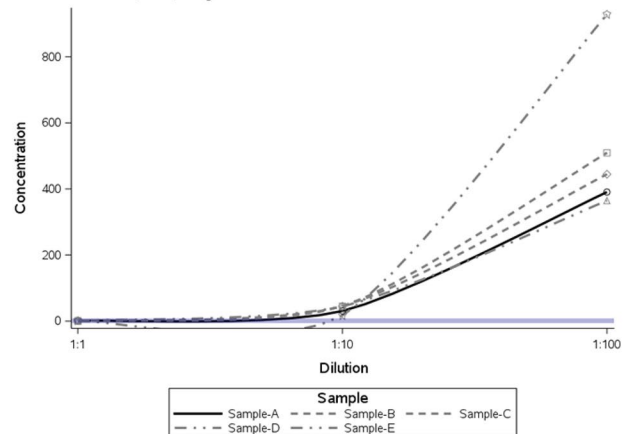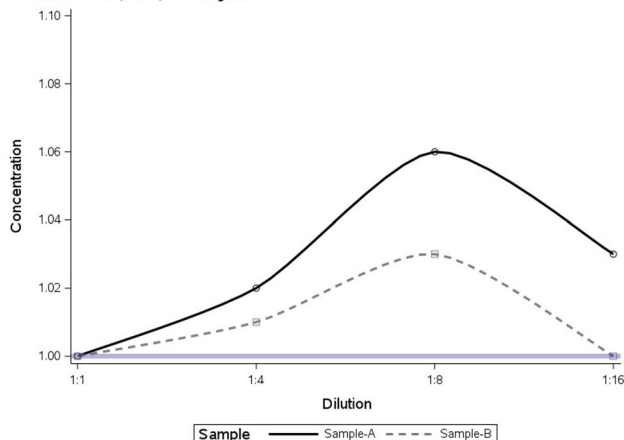
Figure 13. Partial Parallelism Plot for Aβ oligomer quantification. This figure presents the partial parallelism plot for the quantification of Aβ in CSF and PBS [62]. Experiments were conducted with increasing iterations: Sample A had zero iterations, Sample B had one iteration, progressing to Sample E with four iterations. The raw data were kindly shared by the corresponding author. The plots demonstrate a lack of partial parallelism across the dilution range tested. The full dilution range is displayed in the plots on the left. The zoomed-in plots on the right highlight that Aβ oligomers are overestimated at dilutions beyond 1:10.

Figure 14. Partial Parallelism Plot for amyloid β_{1-42} and amyloid β_{1-40} quantified from artificial samples of recombinant human serum albumin, which was spiked with full-length recombinant amyloid β_{1-42} and amyloid β_{1-40} proteins at concentrations about three times above the highest standard. The data were taken from table 6 in reference [65]. The plots do not demonstrate partial parallelism.

Figure 15. Partial Parallelism plot for total Tau spiked into PBS. The data were taken from Table 5 in reference [77]. There is lack of partial parallelism.

[100]. This raises uncertainty about whether these assays would demonstrate partial parallelism under standard laboratory conditions [9].

These findings underscore limitations in the robustness of assay performance and variability in partial parallelism across studies [9,20,23]. Further discussion of these results, including potential sources of bias, is presented in the subsequent section.

5.6. Risk of bias across studies

The primary risk of bias across studies pertains to deviations from established guidelines for testing parallelism [20]. Current recommendations stipulate that parallelism should be demonstrated using a serially diluted sample response curve in the same buffer used for the standard. However, dilution of one pool of

Figure 16. Partial Parallelism plot for spiked CSF $A\beta_{1-40}$ (78,300 ng/L) diluted in an assay buffer. The data were taken from the top left graph in Figure 2 in reference [69]. The plots do not demonstrate partial parallelism.

samples into another pool, as also performed [100], falls outside these guidelines and risks overestimating parallelism.

In seven studies, generating reliable partial parallelism plots from the provided data was not possible [16,70,73,75,76,90,96]. Several limitations contributed to this, including hidden biases that were not pre-specified in the review protocol published on the PROSPERO Registry. One notable example involved the assessment of parallelism for Amyloid-β immunoassays using spiked calibrators. This analysis covered a heterogeneous range of dilution curves across six laboratories employing one or more of seven assays [76]. Parallelism was reported as mean percentages per laboratory, with results ranging from 74% to 344%, indicating significant variability and inconsistency.

Further complicating the assessment, some studies failed to provide key methodological details, such as sample preparation protocols or calibration standards, which may have introduced additional biases to what was reviewed in Table 3. Additionally, variations in experimental design, such as the use of nonstandard dilution matrices or unverified spiking procedures, may have further reduced the comparability and reproducibility of parallelism testing. This underscores the need for stricter adherence to standardized guidelines and more transparent reporting of methodological details to minimize bias and improve the reliability of partial parallelism assessments. Our analysis revealed substantial variability in parallelism assessment. The implications of these findings are explored in the discussion section.
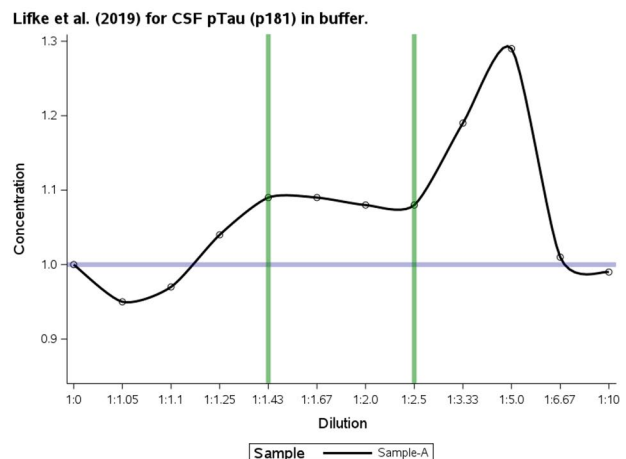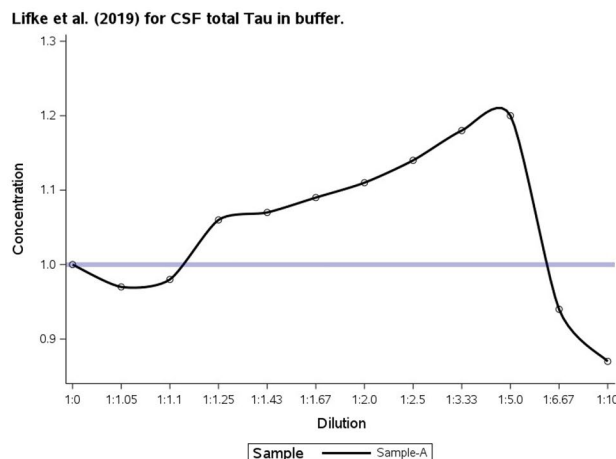
Figure 17. Partial parallelism plots for total Tau and pTau (181 P) assays. Data were derived from "representative results from one sample" in Supplementary Figure S1 of reference [71]. The CSF sample dilution series in buffer was estimated from the x-axis (1, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.15, 0.10), and measured concentrations were approximated from the y-axis. Partial parallelism was achieved within a dilution range of 1:1.43 to 1:2.5 for the pTau181 assay but not for the total Tau assay.

6. Discussion

This systematic review provides a comprehensive analysis of parallelism testing in neurodegeneration biomarker assays. The principal finding is that all current biomarker tests exhibit only partial [23], rather than full, parallelism. Partial parallelism plots, which can be readily generated from existing data, offer a straightforward visual tool for comparison. However, the range of partial parallelism is typically narrow, spanning approximately three doubling dilution steps. This finding has critical implications for interpreting studies that dilute samples beyond this range, as such practices can lead to inaccuracies. Biomarker concentrations may be overestimated when partial parallelism plots deviate

Figure 18. Partial Parallelism plot for $A\beta_{1-42}$ from plasma diluted into a sample buffer. Data were taken from the log-scaled (0–100) y-axis from Figure 3A in reference [72]. The plot shows partial parallelism within a dilution range of 1:4–1:16.

upwards (e.g. Figure 12) or underestimated with downward deviations (e.g. Figure 15). The clinical context of such inaccuracies has been discussed for one example where neurofilaments were quantified at a dilution of 1:400 instead of 1:4. As seen in the present review, other biomarkers used in the context of neurodegeneration are affected as well. Notably, however, near-perfect partial parallelism [9] can sometimes be achieved on a group level (e.g. Figure 10).

The distinction between group-level and individual-level partial parallelism (see Figure 10) highlights two key points. First, biomarker assays showing group-level partial parallelism may be suitable for clinical trials, especially as regulatory agencies like the FDA and EMA increasingly accept biomarker-based endpoints for rapid drug approvals in neurodegeneration. Second, the absence of parallelism in individual samples warrants further investigation. Tests measuring proteins involved in aggregation-related pathologies may harbor hidden biases, which could influence results and interpretations [138]. The effect of age will be one of the most obvious demographic factors to be investigated further [21,139]. Third, partial agreement between different assays for quantification of NfL implies the presence of partial parallelism between the two tests [140]. However, the same method comparison study also showed overestimation of NfL levels by one assay compared to another at higher concentrations [140], indicating a lack of parallelism outside a narrow dilution range. This is entirely consistent with the clinical example, NfL in GBS, discussed earlier.

A 20% failure rate in partial parallelism at the individual sample level indicates that one in five samples may not yield analytically comparable results. This degree of non-parallelism, if attributable to biological
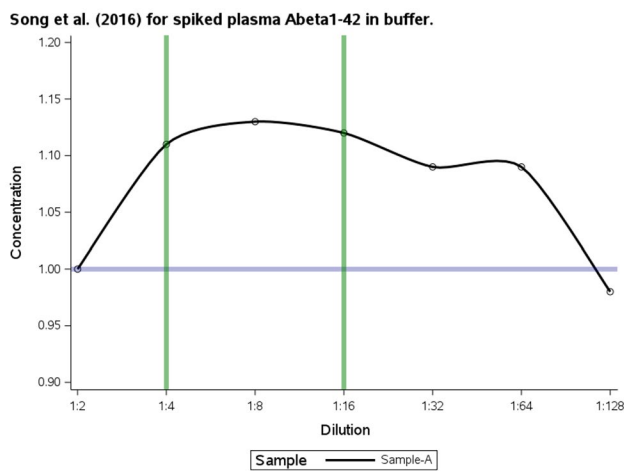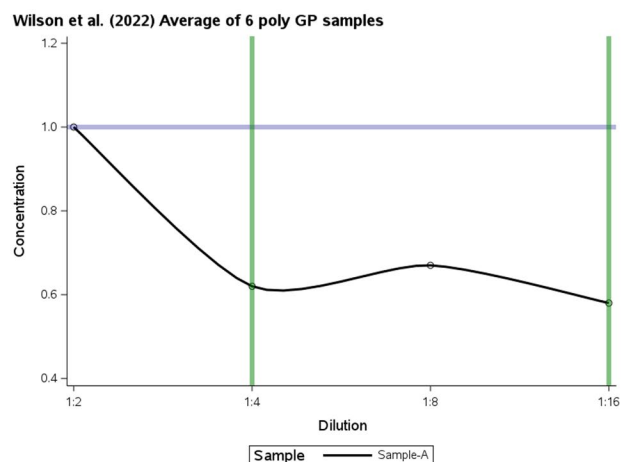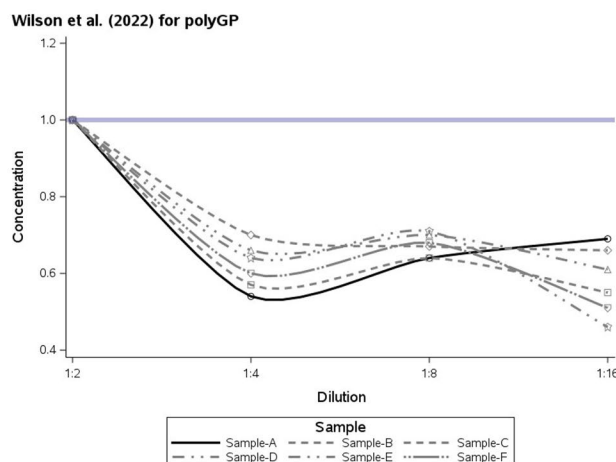

Figure 19. Partial Parallelism plots for quantification of poly-glutamin expansions from native plasma samples. The data were taken from Figure 3H in reference [100]. The plot to the left shows the individual dilution curves. The graph to the right shows the average for 6 plasma samples. On a group level, the plot shows partial parallelism within a dilution range of 1:4–1:16.
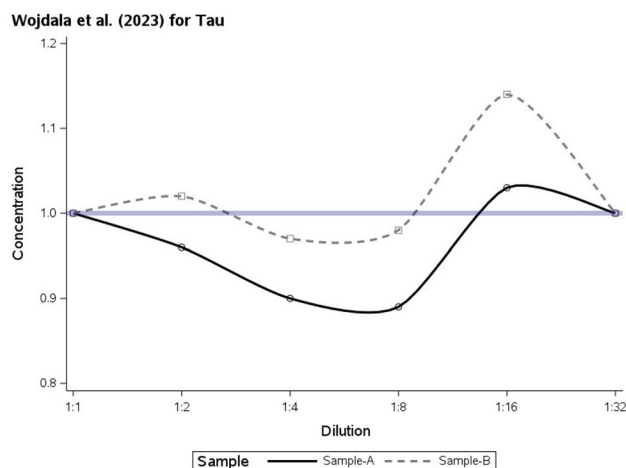
Wojdala et al. (2023) for Tau

Figure 20. Partial Parallelism plot for quantification of phosphorylated (pTau181, pTau231) tau protein from samples spiked with the peptide standard. The data were taken from Supplementary Table S3 in reference [99]. Partial parallelism is not achieved in this plot.

phenomena such as protein aggregation inherent to disease pathology [18,141], could introduce significant bias into quantitative interpretations. It is therefore essential that deviations from parallelism are not dismissed as technical artifacts but are instead rigorously interrogated. While matrix effects are frequently cited, they do not account for all scenarios and are predominantly a concern in mass spectrometry-based platforms [11–13]. Additional causes include endogenous protein–protein interactions, biochemical dissimilarity between native analytes and recombinant standards, analyte instability, suboptimal assay sensitivity, and supraphysiological biomarker concentrations exceeding the assay's dynamic range [18,141,142,11,12,13].

Therefore, a critical finding in this systematic review is the potential publication bias, as study abstracts often overstate the presence of parallelism. Only 14% of the included studies demonstrated clear evidence of partial parallelism [5,61,71,72,100,101]. Interestingly, several studies reported on parallelism between different samples instead of between samples and the standard. This point could be emphasized strongly in future guidelines on harmonization of assay validation and the implementation of quality control.

The review encountered several limitations stemming from inconsistencies in the reporting of methodological details. Dilution ranges varied widely across studies, with arbitrary steps (e.g. 1:4, 1:5, 1:6 versus 1:1,250, 1:2,500, 1:5,000) [69]. While some dilution ranges could be inferred retrospectively [71], this was often based on a single representative sample, preventing the generation of parallelism plots for other data. Additionally, some reported values exceeded the

assay measurement range [71], indicating inappropriate extrapolation beyond the standard curve. For the standard curve, the dilution data were frequently not given. Instead, parallelism between samples, rather than between samples and the standard curve, was shown. Table 3 further highlights that not all samples were diluted in the assay buffer, a critical limitation when testing parallelism with a standard curve [9].

Overall, the number of samples and standard curves was too small for meaningful statistical modeling [42,43]. There is a need for larger numbers [143] of samples, standards and datapoints (of the dilution ranges) for robust statistical evidence [143,144]. For example, one study that was able to demonstrate partial parallelism did so for a very narrow dilution range of 1:1.43–1:2.5 based on one single sample that was considered to be representative [71]. The only study with a very large range of partial parallelism 1:100–1:10,000 reports this close to the detection limit (16 fM) of the test [67].

Moreover, there was a concerning use of spiked or artificial samples in many studies [5,35,61,62,65–69,72,74,77,99,100], instead of the recommended use of native samples [7,8]. For one study, it remained unclear if samples were spiked or not [71]. Taken together, the use of spiked samples reduces the generalizability of the findings for creating representative parallelism plots. Similar issues have been reported for other groups of biomarkers [145]. Stronger adherence to established guidelines is recommended [7].

Likewise, generalizability is hampered by the lack of sharing data on the standard curve for the dilution steps presented [61,35,62,63,64,65,67,68,69,5,71,72,74,100,99,98,77]. On revision of the Figures in the present systematic review, it is not possible to state with absolute certainty if the standard curve overlays the (blue) line of unity. It would be desirable to have, in future studies, data from a representative number of stand curves available. These data points do not need to be exactly for the same dilution steps as for samples if the visual representation with partial parallelism plots is chosen. This is another advantage of the visual approach compared to statistical techniques.

On critical review of the setup for sample dilution, it should be mentioned that, for example, one of the finally excluded studies [16] diluted blood samples from patients with multiple sclerosis into blood samples of healthy controls. For the assessment of parallelism between patient samples and the assay's standard curve, it is mandatory to use the same buffer solution [9,20,21].

Conflicting parallelism results were found for the pTau181 assay. One study did show partial parallelism [71]
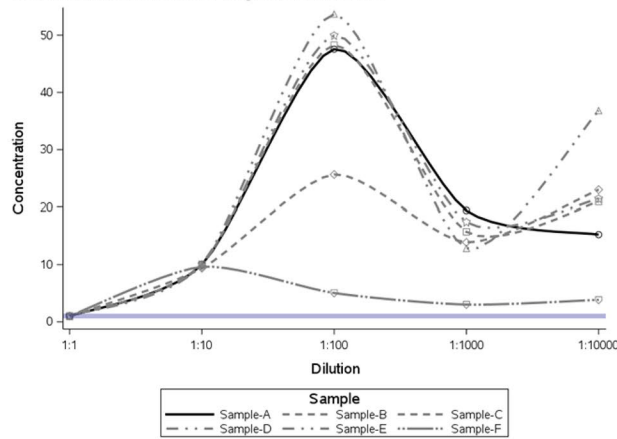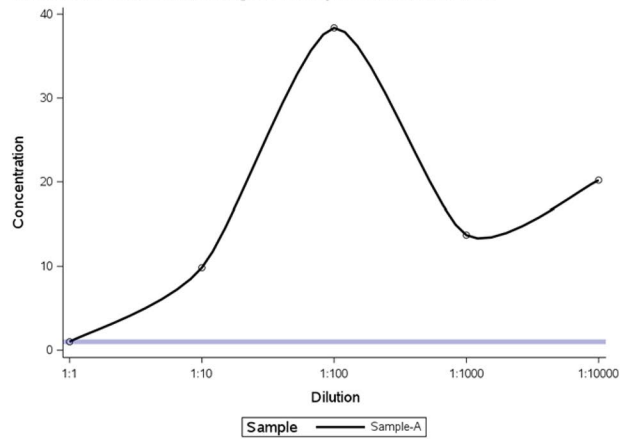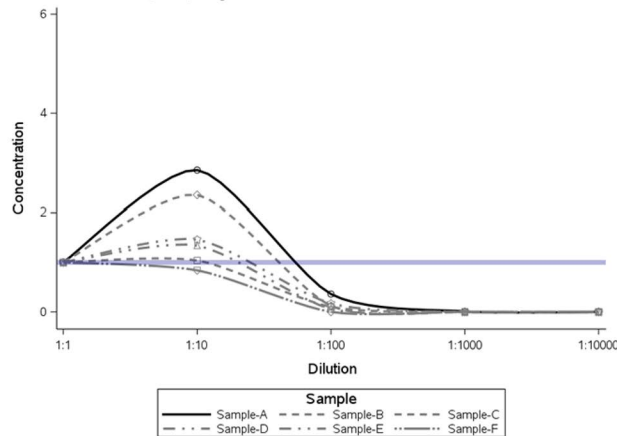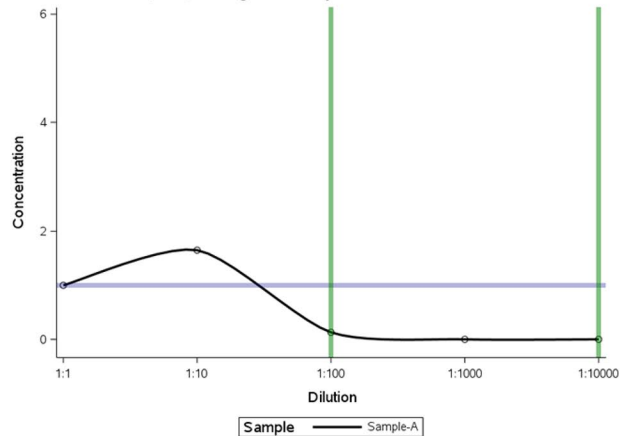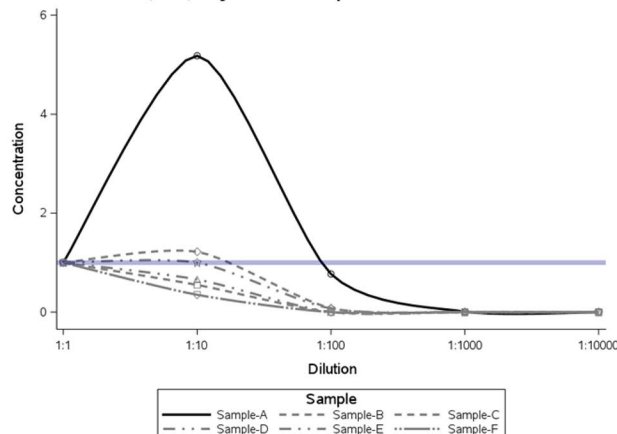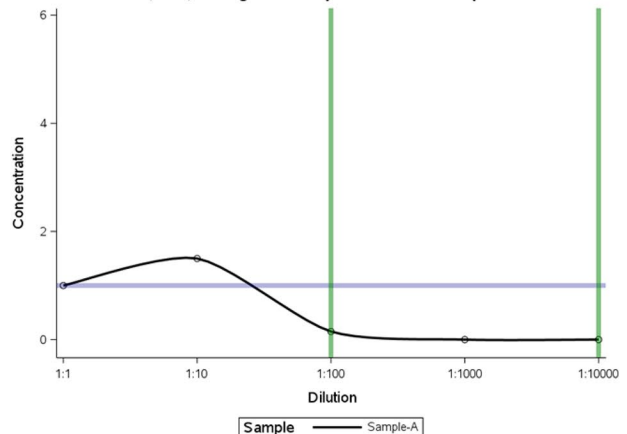
Figure 21. Partial Parallelism Plots for Aβ oligomer quantification in different sampling buffers. The figure shows partial parallelism plots for Aβ oligomers in EDTA, Citrate, and Heparin buffers. Samples A–E are spiked plasma, and Sample F is spiked PBS [67]. The raw data were kindly shared by the corresponding author. The plots reveal a lack of parallelism for EDTA, while partial parallelism is achieved in Citrate and Heparin buffers between dilutions of 1:100 and 1:10,000 on a group level.

(Figure 17B) and one did not [101] (Figures 5A-F). Clearly, on direct visual comparison of these partial parallelism plots, it is evident that different dilution ranges were used. The short stretch of partial parallelism (1:1.43–1:1.25) from one study [71] is for a dilution range lower than the one shown for the other (1:4–1:32 and 1:5–1:40) [101]. Because a third study also did not demonstrate partial parallelism

for pTau181 [99], a balanced interpretation would suggest the absence of parallelism for this biomarker.

Another limitation of this systematic review is the restriction to studies published between 2010 and 2024, a subset of the 32 years since the formulation of the amyloid cascade hypothesis (1992–2024) [102]. This time frame was chosen intentionally, as concerns

regarding the effect of protein aggregation on dilutional parallelism in immunoassays were first raised in 2010 [18], and only subsequently acknowledged in influential white papers [12,21,146] and regulatory guidance documents [20,34,147].

Taken together, this critical and systematic review emphasizes the importance of adhering rigorously to guidelines for parallelism testing [12,20,21,34,146]. These guidelines are embedded in a well-defined framework for laboratory test validation that is endorsed by regulatory authorities (Table 1). If observed, the reason for the lack of partial parallelism needs to be explained, as it has been done in the context of protein aggregation [18]. Authors should explicitly report the dilution range over which partial parallelism is achieved and highlight potential biases, such as over- or underestimation of biomarker concentrations, when dilutions exceed this range.

For clinical trials, it is crucial to clearly document the range of partial parallelism, the dilutions performed, and strategies to minimize biases. Regulatory agencies might consider mandating the inclusion of data from random individual samples to demonstrate parallelism at both individual and group levels. Without clear evidence of partial parallelism, it would be inadvisable to quantify biomarkers for neurodegeneration at varying dilution steps or across different time points in clinical trials. In summary, our critical and systematic review highlights both progress and persistent challenges in parallelism testing. These insights inform recommendations for future assay validation and inclusion into the test validation framework of regulatory authorities (e.g. under precision and bias or under linearity in Table 1).

7. Conclusion

This systematic review identified partial parallelism in only a small proportion of biomarker tests for neurodegeneration. A likely biological reason is the presence of protein aggregates, a key pathological feature in many neurodegenerative diseases. Where partial parallelism is absent, the data suggest specific dilution ranges where it could potentially be achieved. To advance the field, future research must align with established guidelines [20,147], ensuring transparent reporting that allows independent researchers and regulatory bodies to evaluate partial parallelism through standardized plots.

References

[1] Abdelhak A, Foschi M, Abu-Rumeileh S, et al. Blood GFAP as an emerging biomarker in brain and spinal cord disorders. Nat Rev Neurol. 2022;18(3):158–172. doi: 10.1038/s41582-021-00616-3.

[2] Hauser SL, Bar-Or A, Cohen JA, et al. Ofatumumab versus teriflunomide in multiple sclerosis. N Engl J Med. 2020;383(6):546–557. doi: 10.1056/NEJMoa1917246.

[3] Miller TM, Cudkowicz ME, Genge A, et al. Trial of antisense oligonucleotide tofersen for *SOD1* ALS. N Engl J Med. 2022;387(12):1099–1110. doi: 10.1056/NEJMoa2204705.

[4] Khalil M, Teunissen CE, Lehmann S, et al. Neurofilaments as biomarkers in neurological disorders—towards clinical application. Nat Rev Neurol. 2024;20(5):269–287. doi: 10.1038/s41582-024-00955-x.

[5] Lee S, Plavina T, Singh CM, et al. Development of a highly sensitive neurofilament light chain assay on an automated immunoassay platform. Front Neurol. 2022;13:935382. doi: 10.3389/fneur.2022.935382.

[6] Petzold A. The 2022 Lady Estelle Wolfson lectureship on neurofilaments. J Neurochem. 2022;163(3):179–219. doi: 10.1111/jnc.15682.

[7] Andreasson U, Perret-Liaudet A, van Waalwijk van Doorn LJC, et al. A practical guide to immunoassay method validation. Front Neurol. 2015;6:179. doi: 10.3389/fneur.2015.00179.

[8] Burtis C, Ashwood E, editors. Tietz Textbook of Clinical Chemistry. 2nd ed. Philadelphia, Pennsylvania: WB Saunders Company; 1994.

[9] Pum J. A practical guide to validation and verification of analytical methods in the clinical laboratory. Adv Clin Chem. 2019;38:215–281.

[10] Zhang J, Dasgupta A, Ayala R, et al. Internal standard variability: root cause investigation, parallelism for evaluating trackability and practical considerations. Bioanalysis. 2024;16(15):771–776. doi: 10.1080/17576180.2024. 2348939.

[11] Richards S, Amaravadi L, Pillutla R, et al. 2016 White Paper on recent issues in bioanalysis: focus on biomarker assay validation (BAV): (Part 3–LBA, biomarkers and immunogenicity). Bioanalysis. 2016;8(23):2475–2496. doi: 10.4155/bio-2016-4989.

[12] Piccoli S, Mehta D, Vitaliti A, et al. 2019 White paper on recent issues in bioanalysis: FDA immunogenicity guidance, gene therapy, critical reagents, biomarkers and flow cytometry validation (Part 3—recommendations on 2019 FDA immunogenicity guidance, gene therapy bioanalytical challenges, strategies for critical reagent management, biomarker assay validation, flow cytometry validation & CLSI H62). Bioanalysis. 2019;11(24):2207–2244. doi: 10.4155/bio-2019-0271.

[13] Fern'andez-Metzler C, Ackermann B, Garofolo F, et al. Biomarker assay validation by mass spectrometry. Aaps J. 2022;24(3):66.

[14] Petzold A, Altintas A, Andreoni L, et al. Neurofilament ELISA validation. J Immunol Methods. 2010;352(1–2):23–31. doi: 10.1016/j.jim.2009.09.014.

[15] Oeckl P, Jardel C, Salachas F, et al. Multicenter validation of CSF neurofilaments as diagnostic biomarkers for ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2016;17(5–6):404–413. doi: 10.3109/21678421. 2016.1167913.

[16] Wilson D, Chan D, Chang L, et al. Development and multi-center validation of a fully automated digital immunoassay for neurofilament light chain: toward a clinical blood test for neuronal injury. Clin Chem Lab Med. 2024;62(2):322–331. doi: 10.1515/cclm-2023-0518.

[17] Booth RA, Beriault D, Schneider R, et al. Validation and generation of age-specific reference intervals for a new blood neurofilament light chain assay. Clin Chim Acta. 2025;577:120447. doi: 10.1016/j.cca.2025.120447.

[18] Lu CH, Kalmar B, Malaspina A, et al. A method to solubilise protein aggregates for immunoassay quantification which overcomes the neurofilament hook effect. J Neurosci Methods. 2011;195(2):143–150. doi: 10.1016/j. jneumeth.2010.11.026.

[19] Musso G, Zoccarato M, Gallo N, et al. Hook-effect in maglumi immunoassay for serum anti-gad antibodies in neurological disorders: when "wrong" matrix is the right choice. Clin Chim Acta. 2024;558:119679. doi: 10.1016/j.cca.2024.119679.

[20] Committee for Medicinal Products for Human Use. ICH guideline M10 on bioanalytical method validation and study sample analysis. European Medical Agency; 2023 Jan; Available from: https://www.ema.europa.eu/en/ documents/scientific-guideline/ich-guideline-m10-bioanalytical-method-validation-step-5_en.pdf.

[21] Hersey S, Keller S, Mathews J, et al. white paper on recent issues in bioanalysis: ISR for biomarkers, liquid biopsies, spectral cytometry, inhalation/oral & multi-specific biotherapeutics, accuracy/LLOQ for flow cytometry (part 2—recommendations on biomarkers/ CDx assays development & validation, cytometry validation & innovation, biotherapeutics PK LBA regulated bioanalysis, critical reagents & positive controls generation). Bioanalysis. 2021:627–692. 10.4155/bio-2022-0080

[22] Jucker M, Walker LC. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. Nature. 2013;501(7465):45–51. doi: 10.1038/nature12481.

[23] Petzold A. Partial parallelism plots. APPS Applied Sciences. 2024 Jan; 14(2):602. Available from: https:// discovery.ucl.ac.uk/id/eprint/10185313/1/parallel-v9-accepted.pdf.

[24] Winger RM. Gauss and non-euclidean geometry. Bull Amer Math Soc. 1925;31(7):356–358. doi: 10.1090/ S0002-9904-1925-04054-9.

[25] Jenkovszky L, Lake MJ, Soloviev VJ. 'Janos Bolyai, Carl Friedrich Gauss, Nikolai Lobachevsky and the new geometry: foreword. Symmetry (Basel). 2023;15(3):707. doi: 10.3390/sym15030707.

[26] Plikaytis BD, Holder PF, Pais LB, et al. Determination of parallelism and nonparallelism in bioassay dilution curves. J Clin Microbiol. 1994;32(10):2441–2447. doi: 10.1128/jcm.32.10.2441-2447.1994.

[27] Meijering E. A chronology of interpolation: from ancient astronomy to modern signal and image processing. Proc IEEE. 2002;90(3):319–342. doi: 10.1109/5.993400.

[28] Rodbard D, Hutt D. Statistical analysis of radioimmunoassays and immunoradiometric (labelled antibody) assays. a generalized weighted, iterative, least-squares method for logistic curve fitting. NSA-31-000958. 1974:165–192. OSTI ID:4277004.

[29] Rodbard D, Munson P, De Lean A. Improved curve-fitting, parallelism testing, characterization of sensitivity and specificity, validation, and optimization for radioligand assays. Radioimmun Relat Proced Med. 1978;165–192.

[30] Yang H. Emerging non-clinical biostatistics in biopharmaceutical development and manufacturing. Boca Raton, Florida, USA: Chapman and Hall/CRC; 2016.

[31] Prentice RL. A generalization of the probit and logit methods for dose response curves. Biometrics. 1976;32(4):761–768. doi: 10.2307/2529262.

[32] Dudley RA, Edwards P, Ekins RP, et al. Guidelines for immunoassay data processing. Clin Chem. 1985;31(8):1264–1271. doi: 10.1093/clinchem/31.8.1264.

[33] Gottschalk PG, Dunn JR. Determining the error of dose estimates and minimum and maximum acceptable concentrations from assays with nonlinear dose-response curves. Comput Methods Programs Biomed. 2005;80(3): 204–215. doi: 10.1016/j.cmpb.2005.08.003.

[34] Food and Drug Administration. M10 Bioanalytical Method Validation and Study Sample Analysis. 2023;2023.

[35] Cullen VC, Fredenburg RA, Evans C, et al. Development and advanced validation of an optimized method for the

quantitation of aβ42 in human cerebrospinal fluid. Aaps J. 2012;14(3):510–518. doi: 10.1208/s12248-012-9360-7.

[36] Fu Y, Barkley D, Li W, et al. Evaluation, identification and impact assessment of abnormal internal standard response variability in regulated lcms bioanalysis. Bioanalysis. 2020;12(8):545–559. doi: 10.4155/bio-2020-0058.

[37] J Woolf E. Learning how to interpret 'dangerous' internal standard behaviors. BioanalYsis. 2019;11(18):1679–1684. doi: 10.4155/bio-2019-0005.

[38] Lawson NS, Haven GT, Williams GW. Analyte stability in clinical chemistry quality control materials. Crit Rev Clin Lab Sci. 1982;17(1):1–50. doi: 10.3109/10408368209107031.

[39] Roberts CJ. Protein aggregation and its impact on product quality. Curr Opin Biotechnol. 2014;30:211–217. doi: 10.1016/j.copbio.2014.08.001.

[40] Tu J, Bennett P. Parallelism experiments to evaluate matrix effects, selectivity and sensitivity in ligand-binding assay method development: pros and cons. Bioanalysis. 2017;9(14):1107–1122. doi: 10.4155/bio-2017-0084.

[41] Loh TP, Lim CY, Sethi SK, et al. Advances in internal quality control. Crit Rev Clin Lab Sci. 2023;60(7):502–517. doi: 10.1080/10408363.2023.2209174.

[42] Faya P, Rauk AP, Griffiths KL, et al. A curve similarity approach to parallelism testing in bioassay. J Biopharm Stat. 2020;30(4):721–733. doi: 10.1080/10543406.2020.1730875.

[43] Gottschalk PG, Dunn JR. Measuring parallelism, linearity, and relative potency in bioassay and immunoassay data. J Biopharm Stat. 2005;15(3):437–463. doi: 10.1081/BIP-200056532.

[44] Guy S, Shepherd MF, Bowyer AE, et al. How to assess parallelism in factor assays: coefficient of variation of results with different dilutions or slope ratio? Int J Lab Hematol. 2023;45(2):229–240. doi: 10.1111/ijlh.14002.

[45] Jonkman JN, Sidik K. Equivalence testing for parallelism in the four-parameter logistic model. J Biopharm Stat. 2009;19(5):818–837. doi: 10.1080/10543400903105299.

[46] Smith WC, Sittampalam GS. Conceptual and statistical issues in the validation of analytic dilution assays for pharmaceutical applications. J Biopharm Stat. 1998;8(4):509–532. doi: 10.1080/10543409808835257.

[47] Giltinan D. Selected topics in parallel line bioassay: potency estimation by parallel line biossay techniques. Dev Biol (Basel). 2002;107:47–56.

[48] Klein J, Capen R, Mancinelli R, et al. Validation of assays for use with combination vaccines. Biologicals. 1999;27(1):35–41. doi: 10.1006/biol.1999.0167.

[49] Das G. R. Session ii potency estimation by parallel line bioassay techniquesstatistical considerations in the design and analysis of parallel line assays. Dev Biol (Basel). 2002;107:35–46.

[50] Mart'ın-Aguilar L, Camps-Renom P, Lleix A C, et al. Serum neurofilament light chain predicts long-term prognosis in Guillain-Barr'e syndrome patients. Journal of Neurology. Neurosurgery & Psychiatry. 2020;92(1):70–77. '

[51] Axelsson M, Sjogren M, Andersen O, et al. Neurofilament light protein levels in cerebrospinal fluid predict long-term disability of Guillain-Barr'e syndrome: a pilot study. Acta Neurol Scand. 2018;138(2):143–150. doi: 10.1111/ane.12927.

[52] Petzold A, Hinds N, Murray NF, et al. CSF neurofilament levels: a potential prognostic marker in Guillain-Barre'e syndrome. Neurology. 2006;67(6):1071–1073. doi: 10.1212/01.wnl.0000237334.69665.92.

[53] Petzold A, Brettschneider J, Jin K, et al. CSF protein biomarkers for proximal axonal damage improve prognostic accuracy in the acute phase of Guillain-Barr'e syndrome. Muscle Nerve. 2009;40(1):42–49. doi: 10.1002/mus.21239.[]

[54] Yuki N, Hartung HP. Guillain-Barr'e syndrome. N Engl J Med. 2012;366(24):2294–2304. doi: 10.1056/NEJMra1114525.

[55] Guillain G, Barr'e J, Strohl A. Sur un syndrome de radiculo–n'evrite avec hyperalbuminose du liquide c'ephalo–rachidien sans r'eaction cellulaire. Remarques sur les caract'eres cliniques et graphiques des r'eflexes tendineux. Bull Soc M'ed Hˆosp Paris. 1916;40:1462–1470.

[56] Papanicolas I, Rajan D, Karanikolos M, et al., editors. Health system performance assessment: a framework for policy analysis. Geneva: world Health Organization; 2022. Health Policy Series. No. 57; Available from: https://www.who.int/publications/i/item/9789240042476.

[57] van Tilburg SJ, Teunissen CE, Maas CC, et al. Dynamics and prognostic value of serum neurofilament light chain in Guillain-Barr'e syndrome. EBioMedicine. 2024;102:105072. doi: 10.1016/j.ebiom.2024.105072.

[58] Hadden RDM, Cornblath DR, Hughes RAC, et al. Electrophysiological classification of Guillain-Barr'e syndrome: clinical associations and outcome. Plasma Exchange/Sandoglobulin Guillain-Barr'e Syndrome Trial Group. Ann Neurol. 1998;44(5):780–788. doi: 10.1002/ana.410440512.

[59] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi: 10.1136/bmj.n71.

[60] Haddaway NR, Page MJ, Pritchard CC, et al. PRISMA2020: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Sys Rev. 2022;18(2):e1230.

[61] Button EB, Cheng WH, Barron C, et al. Development of a novel, sensitive translational immunoassay to detect plasma glial fibrillary acidic protein (GFAP) after murine traumatic brain injury. Alzheimers Res Ther. 2021;13(1):58.

[62] Herrmann Y, Kulawik A, Kühbach K, et al. sFIDA automation yields sub-femtomolar limit of detection for Aβ aggregates in body fluids. Clin Biochem. 2017;50(4–5):244–247. doi: 10.1016/j.clinbiochem.2016.11.001.

[63] Huynh HH, Kuch K, Orquillas A, et al. Metrologically traceable quantification of 3 apolipoprotein E isoforms in cerebrospinal fluid. Clin Chem. 2023;69(7):734–745. doi: 10.1093/clinchem/hvad056.

[64] Kaneko N, Yamamoto R, Sato TA, et al. Identification and quantification of amyloid beta-related peptides in human plasma using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Proc Jpn Acad Ser B Phys Biol Sci. 2014;90:104–117.

[65] Kirmess KM, Meyer MR, Holubasch MS, et al. The PrecivityAD™ test: accurate and reliable LC-MS/MS as-

says for quantifying plasma amyloid beta 40 and 42 and apolipoprotein e proteotype for the assessment of brain amyloidosis. Clin Chim Acta. 2021;519:267–275. doi: 10.1016/j.cca.2021.05.011.

[66] Koel-Simmelink MJA, Vennegoor A, Killestein J, et al. 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. 2014;402(1–2):43–49. 49. doi: 10.1016/j.jim.2013.11.008.

[67] Kravchenko K, Kulawik A, Hülsemann M, et al. Analysis of anticoagulants for bloodbased quantitation of amyloid β oligomers in the sFIDA assay. Biol Chem. 2017;398(4):465–475. doi: 10.1515/hsz-2016-0153.

[68] Kruse N, Schlossmacher MG, Schulz-Schaeffer WJ, et al. A First Tetraplex Assay for the Simultaneous Quantification of Total α-Synuclein, Tau, β-Amyloid42 and DJ-1 in Human Cerebrospinal Fluid. PLoS One. 2016;11(4):e0153564. doi: 10.1371/journal.pone.0153564.

[69] Lachno DR, Evert BA, Vanderstichele H, et al. Validation of assays for measurement of amyloid-β peptides in cerebrospinal fluid and plasma specimens from patients with Alzheimer's disease treated with solanezumab. J Alzheimers Dis. 2013;34(4):897–910. doi: 10.3233/JAD-122317.

[70] Lachno DR, Evert BA, Maloney K, et al. Validation and Clinical Utility of ELISA Methods for Quantification of Amyloid-β of Peptides in Cerebrospinal Fluid Specimens from Alzheimer's Disease Studies. J Alzheimers Dis. 2015;45(2):527–542. doi: 10.3233/JAD-141686.

[71] Lifke V, Kollmorgen G, Manuilova E, et al. Elecsys(R) Total-Tau and Phospho-Tau (181P) CSF assays: analytical performance of the novel, fully automated immunoassays for quantification of tau proteins in human cerebrospinal fluid. Clin Biochem. 2019;72:30–38. doi: 10.1016/j.clinbiochem.2019.05.005.

[72] Song L, Lachno DR, Hanlon D, et al. A digital enzyme-linked immunosorbent assay for ultrasensitive measurement of amyloid-β 1-42 peptide in human plasma with utility for studies of Alzheimer's disease therapeutics. Alzheimers Res Ther. 2016;8:58.

[73] Thijssen EH, Verberk IMW, Vanbrabant J, et al. Highly specific and ultrasensitive plasma test detects Abeta(1-42) and Abeta(1-40) in Alzheimer's disease. Scientific RePorts. 2021;11(1):9736.

[74] Triana-Baltzer G, Van Kolen K, Theunis C, et al. Development and validation of a high sensitivity assay for measuring p217tau in cerebrospinal fluid. J Alzheimers Dis. 2020;77(4):1417–1430. doi: 10.3233/JAD-200463.

[75] van Waalwijk van Doorn LJC, Koel-Simmelink MJ, Haußmann U, et al. Validation of soluble amyloid-β precursor protein assays as diagnostic CSF biomarkers for neurodegenerative diseases. J Neurochem. 2016;137(1):112–121. doi: 10.1111/jnc.13527.

[76] van Waalwijk van Doorn LJC, Kulic L, Koel-Simmelink MJA, et al. Multicenter analytical validation of aβ40 immunoassays. Front Neurol. 2017;8:310. doi: 10.3389/fneur.2017.00310.

[77] Yang SY, Chiu MJ, Chen TF, et al. Analytical performance of reagent for assaying tau protein in human plasma and feasibility study screening neurodegenerative diseases. Sci Rep. 2017;7(1):9304. doi: 10.1038/s41598-017-09009-3.

[78] Chen L, Lin J, Yi J, et al. A tyrosinase-induced fluorescence immunoassay for detection of tau protein using dopamine-functionalized CuInS(2)/ZnS quantum dots. Anal Bioanal Chem. 2019;411(20):5277–5285. doi: 10.1007/s00216-019-01909-9.

[79] Ciou SH, Hsieh AH, Lin YX, et al. Sensitive label-free detection of the biomarker phosphorylated tau-217 protein in Alzheimer's disease using a graphene-based solution-gated field effect transistor. Biosens Bioelectron. 2023;228:115174. doi: 10.1016/j.bios.2023.115174.

[80] De Simone A, Fiori J, Naldi M, et al. Application of an ESI-QTOF method for the detailed characterization of GSK-3β inhibitors. J Pharm Biomed Anal. 2017;144:159–166. doi: 10.1016/j.jpba.2017.02.036.

[81] Delnomdedieu M, Duvvuri S, Li DJ, et al. First-In-Human safety and long-term exposure data for AAB-003 (PF-05236812) and biomarkers after intravenous infusions of escalating doses in patients with mild to moderate Alzheimer's disease. Alzheimers Res Ther. 2016;8(1):12. doi: 10.1186/s13195-016-0177-y.

[82] Farsani ZA, Schmid VJ. Co-localization analysis in fluorescence microscopy via maximum entropy copula. Int J Biostat. 2020;17(1):165–175. doi: 10.1515/ijb-2019-0019.

[83] Galba J, Michalicova A, Parrak V, et al. Quantitative analysis of phenylalanine, tyrosine, tryptophan and kynurenine in rat model for tauopathies by ultra-high performance liquid chromatography with fluorescence and mass spectrometry detection. J Pharm Biomed Anal. 2016;117:85–90. doi: 10.1016/j.jpba.2015.08.026.

[84] Ge L, Tian Y. Fluorescence lifetime imaging of p-tau protein in single neuron with a highly selective fluorescent probe. Anal Chem. 2019;91(5):3294–3301. doi: 10.1021/acs.analchem.8b03992.

[85] Hupert M, Elfgen A, Schartmann E, et al. Development and validation of an UHPLCESI-QTOF-MS method for quantification of the highly hydrophilic amyloid-β oligomer eliminating all-D-enantiomeric peptide RD2 in mouse plasma. J Chromatogr B Analyt Technol Biomed Life Sci. 2018;1073:123–129. doi: 10.1016/j.jchromb.2017.12.009.

[86] Khan ZA, Park S. Aβ-Ni-HRP sandwich assay: a new sensitive colorimetric method for the detection of Aβ (1-40). Talanta. 2022;237:122946. doi: 10.1016/j.talanta.2021.122946.

[87] Khachornsakkul K, Tiangtrong A, Suwannasom A, et al. Distance-based β-amyloid protein detection on PADs for the scanning and subsequent follow-up of Alzheimer's disease in human urine samples. Analyst. 2022;147(4):695–703. doi: 10.1039/d1an01605a.

[88] Kovac A, Somikova Z, Zilka N, et al. Liquid chromatography-tandem mass spectrometry method for determination of panel of neurotransmitters in cerebrospinal fluid from the rat model for tauopathy. Talanta. 2014;119:284–290. doi: 10.1016/j.talanta.2013.10.027.

[89] Lame ME, Chambers EE, Blatnik M. Quantitation of amyloid beta peptides aβ(1-38), aβ(1-40), and aβ(1-42) in human cerebrospinal fluid by ultra-performance liquid chromatography-tandem mass spectrometry. Anal Biochem. 2011;419(2):133–139. doi: 10.1016/j.ab.2011.08.010.

[90] Liu L, Kwak H, Lawton TL, et al. An ultra-sensitive immunoassay detects and quantifies soluble aβ oligo-

mers in human plasma. Alzheimer Dementia. 2022;18(6):1186–1202. doi: 10.1002/alz.12457.

[91] Pei L, Ye Y, Zhao W, et al. A validated UPLC-MS/MS method for quantitative determination of a potent neuroprotective agent Sarsasapogenin-AA13 in rat plasma: application to pharmacokinetic studies. Biomed Chromatogr. 2020;34(3):e4775. doi: 10.1002/bmc.4775.

[92] Perez FP, Bandeira JP, Morisaki JJ, et al. Antenna design and sar analysis on human head phantom simulation for future clinical applications. J Biomed Sci Eng. 2017;10(9):421–430. doi: 10.4236/jbise.2017.109032.

[93] Pils M, Dybala A, Rehn F, et al. Development and Implementation of an Internal Quality Control Sample to Standardize Oligomer-Based Diagnostics of Alzheimer's Disease. Diagnostics. 2023;13(10):1702. doi: 10.3390/diagnostics13101702.

[94] Shapiro GI, LoRusso P, Dowlati A, et al. A Phase 1 study of RO6870810, a novel bromodomain and extra-terminal protein inhibitor, in patients with NUT carcinoma, other solid tumours, or diffuse large B-cell lymphoma. Br J Cancer. 2021;124(4):744–753. doi: 10.1038/s41416-020-01180-1.

[95] Vangavaragu JR, Valasani KR, Fang D, et al. Determination of small molecule ABAD inhibitors crossing blood-brain barrier and pharmacokinetics. J Alzheimers Dis. 2014;42(1):333–344. doi: 10.3233/JAD-140252.

[96] Wang Y, Yu Z, Ning Z, et al. Development of a time-resolved immunochromatographic test strip for rapid and quantitative determination of GFAP in serum. Mikrochim Acta. 2024;191(6):325. doi: 10.1007/s00604-024-06385-2.

[97] Zhang H, Gopinath SCB, Hu Y. Spinal cord injury immunosensor: sensitive detection of S100β on interdigitated electrode sensor. Heliyon. 2023;9(9):e19304. doi: 10.1016/j.heliyon.2023.e19304.

[98] Yamashita K, Watanabe S, Ishiki K, et al. Fully automated chemiluminescence enzyme immunoassays showing high correlation with immunoprecipitation mass spectrometry assays for β-amyloid (1-40) and (1-42) in plasma samples. Biochem Biophys Res Commun. 2021;576:22–26. doi: 10.1016/j.bbrc.2021.08.066.

[99] Wojdal-A AL, Bellomo G, Gaetani L, et al. Targeting multiphosphorylated tau: technical and clinical validation of a new Simoa® assay for CSF and plasma detection of tau simultaneously phosphorylated at T181 and T231. medRxiv. 2023. doi: 10.1101/2023.06.08.23291128

[100] Wilson KM, Katona E, Glaria I, et al. Development of a sensitive trial-ready poly(GP) CSF biomarker assay for C9orf72-associatedfrontotemporal dementia and amyotrophiclateral sclerosis. JNNP. 2022;93(7):761–771.

[101] Bayoumy S, Verberk IMW, den Dulk B, et al. Clinical and analytical comparison of six Simoa assays for plasma P-tau isoforms P-tau181, P-tau217, and P-tau231. Alzheimers Res Ther. 2021;13(1):198.

[102] Hardy J, Higgins G. Alzheimers disease: the amyloid cascade hypothesis. Science. 1992;256(5054):184–185. doi: 10.1126/science.1566067.

[103] Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. 2007;6(8):734–746. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi? cmd=prlinks&dbfrom=pubmed&retmode=ref&id=17616482. doi: 10.1016/S1474-4422(07)70178-3.

[104] Dubois B, Feldman HH, Jacova C, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 2014;13(6):614–629. doi: 10.1016/S1474-4422(14)70090-0.

[105] Dubois B, Villain N, Schneider L, et al. Alzheimer disease as a clinical-biological construct—an international working group recommendation. JAMA Neurol. 2024;81(12):1304–1311. doi: 10.1001/jamaneurol.2024.3770.

[106] Pitschke M, Prior R, Haupt M, et al. Detection of single amyloid beta-protein aggregates in the cerebrospinal fluid of Alzheimer's patients by fluorescence correlation spectroscopy. Nat Med. 1998;4(7):832–834. doi: 10.1038/nm0798-832.

[107] Nasrollah RG, Mehriar A, Karin G, et al. Familial Alzheimer's disease-related mutations differentially alter stability of amyloid-beta aggregates. J Phys Chem Lett. 2023;14(6):1427–1435.

[108] O'Brien RJ, Wong PC. Amyloid precursor protein processing and Alzheimer's disease. Annu Rev Neurosci. 2011;34(1):185–204.

[109] Antony D, Sheth P, Swenson A, et al. Recent advances in Alzheimer's disease therapy: clinical trials and literature review of novel enzyme inhibitors targeting amyloid precursor protein. Expert Opin Pharmacother. 2025;26(1):63–73. doi: 10.1080/14656566.2024.2438317.

[110] Danzer KM, Krebs SK, Wolff M, et al. Seeding induced by α-synuclein oligomers pro-vides evidence for spreading of α-synuclein pathology. J Neurochem. 2009;111(1):192–203. doi: 10.1111/j.1471-4159.2009.06324.x.

[111] Hong Z, Shi M, Chung KA, et al. DJ-1 and alpha-synuclein in human cerebrospinal fluid as biomarkers of Parkinson's disease. Brain. 2010;133(Pt 3):713–726. doi: 10.1093/brain/awq008.

[112] Grozdanov V, Bousset L, Hoffmeister M, et al. Increased immune activation by patho-logic α-Synuclein in Parkinson's disease. Ann Neurol. 2019;86(4):593–606. doi: 10.1002/ana.25557.

[113] Magalhães P, Lashuel HA. Opportunities and challenges of alpha-synuclein as a potential biomarker for Parkinson's disease and other synucleinopathies. NPJ Parkinson's Dis. 2022;8(1). doi:10.1038/s41531-022-00357-0

[114] Uversky VN. Neuropathology, biochemistry, and biophysics of α-synuclein aggregation. J Neurochem. 2007;103(1):17–37. doi: 10.1111/j.1471-4159.2007.04764.x.

[115] Waragai M, Wei J, Fujita M, et al. Increased level of DJ-1 in the cerebrospinal fluids of sporadic Parkinson's disease. Biochem Biophys Res Commun. 2006;345(3):967–972. doi: 10.1016/j.bbrc.2006.05.011.

[116] Blennow K, Hampel H, Weiner M, et al. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol. 2010;6(3):131–144. doi: 10.1038/nrneurol.2010.4.

[117] Ishiguro K, Ohno H, Arai H, et al. Phosphorylated tau in human cerebrospinal fluid is a diagnostic marker for Alzheimer's disease. Neurosci Lett. 1999;270(2):91–94. doi: 10.1016/s0304-3940(99)00476-0.

[118] Sjogren M, Davidsson P, Tullberg M, et al. Both total and phosphorylated tau are increased in Alzheimer's disease.

Journal of Neurology, Neurosurgery and Psychiatry. 2001;70(5):624–630. doi: 10.1136/jnnp.70.5.624.

[119] Arai H, Terajima M, Miura M, et al. Tau in cerebrospinal fluid: a potential diagnostic marker in Alzheimer's disease. Ann Neurol. 1995;38(4):649–652. doi: 10.1002/ana.410380414.

[120] Blennow K, Wallin A, Agren H, et al. Tau protein in cerebrospinal fluid: a biochemical marker for axonal degeneration in Alzheimer disease? Mol Chem Neuropathol. 1995;26(3):231–245. doi: 10.1007/BF02815140.

[121] Vanderstichele, Blennow Heuvaert. HK, D'ND, et al. CSF: progress in Alzheimer's and Parkinson's diseases. Plenum Press; 1998. Chapter Development of a specific diagnostic test for measurement of beta-amyloid (142); p. 773–778. Available from: file:///home/petz1/lit/pdf-files/1998-book-vanderstichele-amyloid-assay-congres-eilat.pdf:PDF.

[122] Petzold A, Chapman MD, Schraen S, et al. An unbiased, staged, multicentre, validation strategy for Alzheimer's disease CSF tau levels. Exp Neurol. 2010;223(2):432–438. doi: 10.1016/j.expneurol.2009.11.010.

[123] Karikari TK, Ashton NJ, Brinkmalm G, et al. Blood phospho-tau in Alzheimer disease: analysis, interpretation, and clinical utility. Nat Rev Neurol. 2022;18(7):400–418. doi: 10.1038/s41582-022-00665-2.

[124] Ashton NJ, Brum WS, Di Molfetta G, et al. Diagnostic accuracy of a plasma phosphorylated Tau 217 immunoassay for Alzheimer disease pathology. JAMA Neurol. 2024;81(3):255–263. doi: 10.1001/jamaneurol.2023.5319.

[125] William Rebeck G, Reiter JS, Strickland DK, et al. Apolipoprotein E in sporadic Alzheimer's disease: allelic variation and receptor interactions. Neuron. 1993;11(4):575–580. doi: 10.1016/0896-6273(93)90070-8.

[126] Hofman A, Ott A, Breteler M, et al. Atherosclerosis, apolipoprotein E, and prevalence of dementia and Alzheimer's disease in the Rotterdam Study. Lancet. 1997;349(9046):151–154. doi: 10.1016/S0140-6736(96)09328-2.

[127] Nicoll J, Roberts G, Graham D. Apolipoprotein E epsilon 4 allele is associated with deposition of amyloid beta-protein following head injury. Nat Med. 1995;1(2):135–137. doi: 10.1038/nm0295-135.

[128] Yaldizli Ö, Benkert P, Achtnichts L, et al. Personalized treatment decision algorithms for the clinical application of serum neurofilament light chain in multiple sclerosis: a modified delphi study. Mult Scler. 2025;31(8):932–943. doi: 10.1177/13524585251335466.

[129] Petzold A, Bowser R, Calabresi P, et al. Biomarker time out. Mult Scler. 2014;20(12):1560–1563. doi: 10.1177/1352458514524999.

[130] Albrechtsen M, Bock E. Quantification of glial fibrillary acidic protein (GFAP) in human body fluids by means of ELISA employing a monoclonal antibody. J Neuroimmunol. 1985;8(4–6):301–309. doi: 10.1016/s0165-5728(85)80069-2.

[131] Petzold A. Glial fibrillary acidic protein is a body fluid biomarker for glial pathology in human disease. Brain Res. 2015;1600:17–31. doi: 10.1016/j.brainres.2014.12.027.

[132] Cong S, Xiang C, Wang H, et al. Diagnostic utility of fluid biomarkers in multiple system atrophy: a systematic review and meta-analysis. J Neurol. 2021;268(8):2703–2712. doi: 10.1007/s00415-020-09781-9.

[133] Andersen AD, Binzer M, Stenager E, et al. Cerebrospinal fluid biomarkers for Parkinson's disease a systematic review. Acta Neurol Scand. 2017;135(1):34–56. doi: 10.1111/ane.12590.

[134] Forgrave LM, Ma M, Best JR, et al. The diagnostic performance of neurofilament light chain in csf and blood for Alzheimer's disease, frontotemporal dementia, and amyotrophic lateral sclerosis: a systematic review and meta-analysis. Alzheimer's dementia. (Amsterdam, Netherlands); 2019. 11:730–743.

[135] Karantali E, Kazis D, McKenna J, et al. Neurofilament light chain in patients with a concussion or head impacts: a systematic review and meta-analysis. Eur J Trauma Emerg Surg. 2022;48(3):1555–1567. doi: 10.1007/s00068-021-01693-1.

[136] Mondello S, Sorinola A, Czeiter E, et al. Blood-based protein biomarkers for the management of traumatic brain injuries in adults presenting to emergency departments with mild brain injury: a living systematic review and meta-analysis. J Neurotrauma. 2021;38(8):1086–1106. doi: 10.1089/neu.2017.5182.

[137] Petzold A, Keir G, Warren J, et al. A systematic review and meta-analysis of CSF neurofilament protein levels as biomarkers in dementia. Neurodegener Dis. 2007;4(2–3):185–194. doi: 10.1159/000101843.

[138] Plebani. M. Errors in clinical laboratories or errors in laboratory medicine? Clin Chem Lab Med. 2006;44(6):750–759.

[139] Musso G, Gabelli C, Puthenparampil M, et al. Blood biomarkers for Alzheimer's disease with the Lumipulse automated platform: age-effect and clinical value interpretation. Clin Chim Acta. 2025;565:120014. doi: 10.1016/j.cca.2024.120014.

[140] Primiano G, Luigetti M, Righi D, et al. Comparative analysis of SiMoA and Ella immunoassay platforms for measuring serum neurofilament light chain levels in ATTRv with polyneuropathy and presymptomatic carriers. Eur J Neurol. 2025;32(6):e70215. doi: 10.1111/ene.70215.

[141] Adiutori R, Aarum J, Zubiri I, et al. The proteome of neurofilament-containing protein aggregates in blood. Biochem Biophys Rep. 2018;14:168–177. doi: 10.1016/j.bbrep.2018.04.010.

[142] Arnold ME, Booth B, King L, et al. Workshop report: crystal City VI—bioanalytical method validation for biomarkers. Aaps J. 2016;18(6):1366–1372. doi: 10.1208/s12248-016-9946-6.

[143] Seneta E. A tricentenary history of the law of large numbers. Bernoulli. 2013;19(4):1088–1121. doi: 10.3150/12-BEJSP12.

[144] Durrett R. Probability: theory and examples. Vol 49. Cambridge, England: Cambridge University Press; 2019.

[145] Banks RE. Measurement of cytokines in clinical samples using immunoassays: problems and pitfalls. Crit Rev Clin Lab Sci. 2000;37(2):131–182. doi: 10.1080/10408360091174187.

[146] Stevenson L, Garofolo F, Desilva B, et al. 2013 White Paper on recent issues in bioanalysis: "hybrid" the best of LBA and LCMS. Bioanalysis. 2013;5(23):2903–2918. doi: 10.4155/bio.13.238.

[147] Food and Drug Administration. Bioanalytical Method Validation for Biomarkers: Guidance for Industry; 2025.

[148] Petzold, Axel, Dataset for Partial Parallelism Plot Analysis in Neurodegeneration Biomarker Assays (2010–2024), Dataset UCL Figshare, 2025, doi: 10.5522/04/29389949.v1.