Received: 4 May 2022 | Revised: 1 August 2022 | Accepted: 11 August 2022

DOI: 10.1002/alz.12801

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Clinical performance and robustness evaluation of plasma amyloid-$\beta_{42/40}$ prescreening

Christina Rabe[1] | **Tobias Bittner**[1,2] | **Alexander Jethwa**[3] | **Ivonne Suridjan**[4] | **Ekaterina Manuilova**[3] | **Michel Friesenhahn**[1] | **Erik Stomrud**[5,6] | **Henrik Zetterberg**[7,8,9,10,11] | **Kaj Blennow**[7,8] | **Oskar Hansson**[5,6] | **for the Alzheimer's Disease Neuroimaging Initiative**† **and the Swedish BioFINDER study**

[1]Genentech, Inc., South San Francisco, California, USA

[2]F. Hoffmann-La Roche Ltd, Basel, Switzerland

[3]Roche Diagnostics GmbH, Mannheim, Germany

[4]Roche Diagnostics International Ltd, Rotkreuz, Switzerland

[5]Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden

[6]Memory Clinic, Skåne University Hospital, Malmö, Sweden

[7]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

[8]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden

[9]Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK

[10]UK Dementia Research Institute at UCL, London, UK

[11]Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China

Correspondence
Christina Rabe, Genentech, Inc., South San Francisco, CA, USA.
Email: rabe.christina@gene.com

†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Funding information
F. Hoffmann-La Roche Ltd, Basel, Switzerland

Abstract

Introduction: Further evidence is needed to support the use of plasma amyloid β (Aβ) biomarkers as Alzheimer's disease prescreening tools. This study evaluated the clinical performance and robustness of plasma Aβ_{42}/Aβ_{40} for amyloid positivity prescreening.

Methods: Data were collected from 333 BioFINDER and 121 Alzheimer's Disease Neuroimaging Initiative study participants. Risk and predictive values versus percentile of plasma Aβ_{42}/Aβ_{40} evaluated the actionability of plasma Aβ_{42}/Aβ_{40}, and simulations modeled the impact of potential uncertainties and biases. Amyloid PET was the brain amyloidosis reference standard.

Results: Elecsys plasma Aβ_{42}/Aβ_{40} could potentially rule out amyloid pathology in populations with low-to-moderate amyloid positivity prevalence. However, simulations showed small measurement or pre-analytical errors in Aβ_{42} and/or Aβ_{40} cause misclassifications, impacting sensitivity or specificity. The minor fold change between amyloid PET positive and negative cases explains the biomarkers low robustness.

Discussion: Implementing plasma $A\beta_{42}/A\beta_{40}$ for routine clinical use may pose significant challenges, with misclassification risks.

KEYWORDS
Alzheimer's disease, amyloid, biomarkers, blood biomarkers, prescreening

Highlights

- Plasma $A\beta_{42}/A\beta_{40}$ ruled out amyloid PET positivity in a setting of low amyloid-positive prevalence.
- Including (pre-) analytical errors or measurement biases caused misclassifications.
- Plasma $A\beta_{42}/A\beta_{40}$ had a low inherent dynamic range, independent of analytical method.
- Other blood biomarkers may be easier to implement as robust prescreening tools.

1 | INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disease associated with memory decline.[1,2] The key neuropathologic features of AD include the accumulation of amyloid β (Aβ) and phosphorylated tau (pTau) aggregates in the form of plaques and tangles that cause downstream neurodegeneration.[1,2] Aβ plaques build up years before symptoms emerge and, as the disease progresses, lead to memory and functional disability.[1]

Amyloid positron emission tomography (PET) and cerebrospinal fluid (CSF) biomarkers are currently used to confirm underlying amyloid pathology in vivo.[3] However, these methods are expensive, have limited availability, or are perceived as invasive.[4] Globally, there is an urgent unmet need for cost-effective and minimally invasive methods to identify patients with amyloid pathology early in their disease.[3,5] Plasma biomarkers are a proposed solution to prescreen patients into clinical trials, as well as routine clinical practice, especially to help meet the increased demand caused by the availability of disease-modifying therapies.[6]

Multiple technologies, whether mass spectrometry or immunoassay-based, show promising technical and clinical performance when measuring the ratio of $A\beta_{42}/A\beta_{40}$ in plasma.[7–10] These evaluations have been performed in retrospective studies that minimize variability (i.e., banked samples collected under a standardized protocol and measured in a single batch) without the need of prespecifying a cutoff (for each study, the optimal cutoff was identified after the biomarker values were available). Two questions must be addressed to understand the value of blood-based biomarkers for AD in clinical trials and routine better: actionability and robustness for prospective use.

We defined the actionability of a biomarker as its ability to enable decisions for a particular intended use. Two intended uses were considered: (1) as a cost-effective prescreening for clinical trials and (2) as a tool to improve diagnostic work-up for AD in clinical routine. The performance measures to evaluate a biomarker should support the intended use(s). Many studies report receiver operating characteristic area under the curve (ROC-AUC), which measures discrimination but is not directly interpretable for a specific intended use.[11,12] A more informative performance evaluation tool for assessing actionability is the predictiveness curve (see Methods section).

Robustness is the second critical challenge because prospective biomarker measurements to enroll patients for a clinical trial, or for use in clinical routine, will inevitably come with higher variability compared with retrospective measurements of banked samples,[13,14] and they require a prespecified cutoff. Biologic variability (e.g., fluctuations of biomarker concentrations within or between days), pre-analytical variability (e.g., differences in sample handling), and analytical variability (e.g., different laboratories, instruments, reagent lots) can only be standardized and controlled to a certain extent.[15,16] Ideally, low levels of variability should not impact the clinical performance of a biomarker. As in other medical fields, every biomarker needs to be assessed based on its total allowable error (TAE) in measurements that can be tolerated without invalidating the medical usefulness of the analytical result.[13] The TAE is composed of both random and systematic error. A prospective, real-world evaluation in a population that reflects the intended use of the biomarker, and all possible components of variability, is not always feasible, especially in proof-of-concept phases. However, there are indirect approaches to assessing biomarker robustness, based on simulations. Requirements for the TAE (due to all sources) should be derived based on its impact on clinical performance.[13,17]

We aimed to evaluate the clinical performance and robustness of plasma $A\beta_{42}/A\beta_{40}$ as a prescreening tool for AD clinical trials and in routine clinical practice. To better illustrate the potential robustness issue for plasma $A\beta_{42}/A\beta_{40}$, we compared the impact of measurement error on plasma $A\beta_{42}/A\beta_{40}$ with two established biomarker ratios in CSF ($A\beta_{42}/A\beta_{40}$ and $pTau181/A\beta_{42}$), using data from the Swedish BioFINDER study. The Foundation for the National Institutes of Health (FNIH) plasma platform comparison study, using samples from the Alzheimer's Disease Neuroimaging Initiative (ADNI)

RESEARCH IN CONTEXT

1. **Systematic Review**: The authors gathered data from 333 participants from the BioFINDER study, and 121 participants from the Alzheimer's Disease Neuroimaging Initiative studies, which were analyzed and complemented with simulation studies.
2. **Interpretation**: Whereas plasma $A\beta_{42}/A\beta_{40}$ showed potential for ruling out amyloid positivity in a setting of low prevalence of amyloid positivity, it lacked feasibility in moderate prevalence settings. A high risk of misclassification occurred with the introduction of small measurement or pre-analytic errors. This lack of robustness is caused by the narrow dynamic range of plasma $A\beta_{42}/A\beta_{40}$, which is inherent to the biomarker and independent of measurement technology.
3. **Future Directions**: These findings demonstrate significant challenges implementing plasma $A\beta_{42}/A\beta_{40}$ as a robust biomarker for either clinical trials or routine clinical practice. Other promising blood biomarkers with higher robustness will likely prove easier to implement in the clinic.

study, was used to compare different plasma $A\beta_{42}/A\beta_{40}$ measurement technologies.[10]

2 | METHODS

2.1 | BioFINDER subset

The prospective, longitudinal BioFINDER study included consecutively recruited participants from Sweden between 2009 and 2015 (NCT01208675). Data from the subset of patients ($n = 333$) classified as either cognitively unimpaired (CU), with subjective cognitive decline (SCD), or mild cognitive impairment (MCI).[18] All patients in this subset have plasma and CSF biomarker measurements, as well as PET scans available to allow for a comparison of all biomarkers on the same data set. See Table S1 for BioFINDER demographics by visual read PET status and overall status. Details on study design and sample collection have been published previously.[9]

2.2 | ADNI subset

ADNI plasma biomarker data from 130 samples from 121 unique patients were downloaded on June 27, 2021 from the Laboratory of NeuroImaging.[19] Table S2 for ADNI FNIH subset demographics by PET status and overall status.

2.3 | Biomarker assays

BioFINDER Elecsys measurements were performed on a cobas e 601 analyzer using the prototype Elecsys $A\beta_{42}$ and $A\beta_{40}$ immunoassays for plasma (modified improved version compared with the previously published version[9]) and CSF $A\beta_{42}$, $A\beta_{40}$, and pTau181 as described previously.[20] Cutoffs for CSF pTau181/$A\beta_{42}$ and CSF $A\beta_{42}/A\beta_{40}$ were 0.022[20] and 0.059,[9] respectively.

Plasma measurements in ADNI were performed using three immunoassays (Elecsys, Quanterix, ADx) and three mass spectrometry (Washington University, Shimadzu, University of Gothenburg) methods, with a different version and standardization for the Elecsys assays compared with BioFINDER (details in Supplementary Materials).[21]

2.4 | Reference: Amyloid PET

Amyloid PET was used as a common reference for all markers (CSF and plasma). For BioFINDER, $A\beta$ positivity was determined by flutemetamol PET visual read as previously described.[20] The ADNI study used florbetapir PET amyloid images; amyloid positivity was defined as standardized uptake value ratio ≥ 1.11, since no visual read was available for these scans.[10]

2.5 | Clinical performance measures

The most commonly reported performance measures when comparing a quantitatively measured biomarker with a binary reference are ROC curves and AUC, as well as sensitivity and specificity of selected cutoffs.[11,12] However, other measures are more relevant from a clinical standpoint[22]: positive predictive value (PPV), the probability that a patient is indeed PET+ if the biomarker test is positive (i.e., PET+ rate in biomarker test +) and the negative predictive value (NPV), the probability that a patient is indeed PET− if the biomarker test is negative (i.e., PET− rate in biomarker test −). Both PPV and NPV depend on the prevalence in the population. These are cumulative probabilities and refer to a group of patients that are below or above a cutoff. The probability of being positive ("risk") at a specific biomarker concentration can be estimated,[12] and is more powerful for clinical decision-making for an individual patient.[12,22] Based on the risk curve, meaningful cutoffs for risk categories for further guidance can be derived.

Population performance, meaning how frequently various risks occur, is important to understand when evaluating a biomarker's usefulness.[12] A biomarker that determines extreme risks in a larger fraction of the intended use population will have more utility.[12] The predictiveness curve is a useful tool introduced by Pepe *et al.*[12] that evaluates the risk distribution by plotting the risk versus the biomarker percentile.

When evaluating a biomarker as a prescreener for a clinical trial, PPV and NPV describe the cumulative risk in the populations screened-out and enrolled. We propose an integrated risk-profile plot showing risk and cumulative risk versus the percentile of the biomarker

distribution (i.e., proportion of patients below threshold), which contains all information necessary to assess the utility of a biomarker (see Figure S1 and Supplementary Materials for details on interpretation, estimation, and implementation).

2.6 | Robustness

The TAE is a quality requirement that sets a limit for combined imprecision (random error) and bias (systematic error) that are tolerable in a single measurement or single test result to ensure clinical usefulness.[13,23] Imprecision (measured as coefficient of variation [CV]) and bias are often combined to a total error (TE) using TE = bias + 1.65*CV, but quality goals need to be derived for both bias and TE.[17,23] Simulation studies (indirect methods) can help to understand the impact of potential measurement uncertainty on clinical outcomes and to derive the TAE and bias goals.[13]

If a cutoff or the associated risk based on a biomarker concentration is determined under certain conditions (e.g., a specific lab, reagent lot, or pre-analytical sample handling), any change in condition that leads to different concentrations needs to be treated as bias. Bias can have a direct and large impact on clinical decisions as they move a group of patients systematically above or below a threshold.[17] While the impact of bias can be easily investigated by simply shifting the distribution, simulation studies can combine bias and imprecision (see details in the Supplementary Materials).

3 | RESULTS

3.1 | BioFINDER clinical performance

Concordance between plasma $A\beta_{42}/A\beta_{40}$ and amyloid PET (AUC 0.793), CSF $A\beta_{42}/A\beta_{40}$ (AUC 0.825), and CSF pTau/$A\beta_{42}$ (AUC 0.813) remained meaningfully lower than concordance between CSF ($A\beta_{42}/A\beta_{40}$ or pTau/$A\beta_{42}$) and PET (AUC 0.942 and 0.956).

Figure 1A shows the estimated risk of being PET+ for each plasma $A\beta_{42}/A\beta_{40}$ value in the study population with the observed prevalence of ~ 30% PET positivity. For high plasma $A\beta_{42}/A\beta_{40}$ levels, the risk of being PET+ was extremely low (i.e., close to zero; note that the x-axis is reversed) and very high risks were not observed (i.e., the highest risk estimate does not exceed 60%).

Figure 1B illustrates the risk versus the percentile of plasma $A\beta_{42}/A\beta_{40}$ levels in the population (note the x-axis shows percentile with respect to proportion in non-pathological direction), that is, the distribution of risk that describes the utility of a biomarker for a population (predictiveness curve). Very low risks occurred in about 25% of the population (25% highest plasma $A\beta_{42}/A\beta_{40}$ levels).

The estimated risk depends on the prevalence in the population; Figure S2 shows how the risk curve would change with a different prevalence. These curves assume constant sensitivity and specificity, that is, the conditional distribution of plasma $A\beta_{42}/A\beta_{40}$ per given PET status is similar, only the prevalence changes. While this is a simplification and makes strong assumptions, Figure S3 provides evidence that this may be applicable for plasma $A\beta_{42}/A\beta_{40}$ across the observed disease stages of CU, SCD, and MCI.

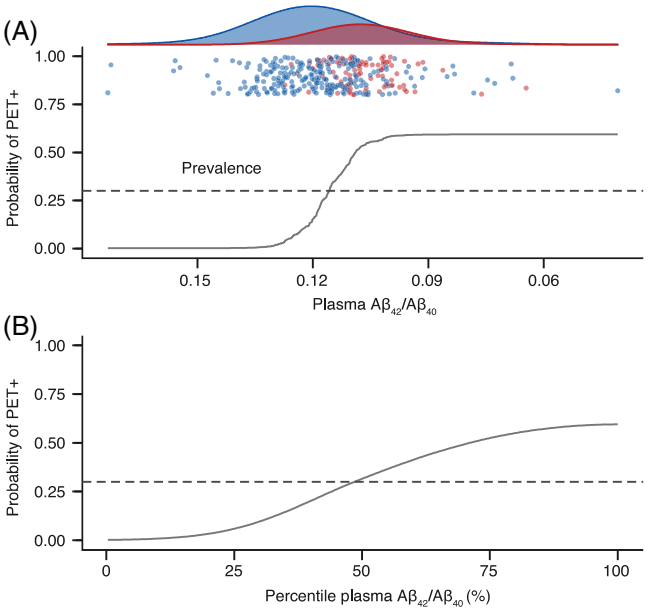

FIGURE 1 Estimated risk of being amyloid PET+ versus (A) plasma $A\beta_{42}/A\beta_{40}$ value[a] and (B) percentile of plasma $A\beta_{42}/A\beta_{40}$[b]. $A\beta$, amyloid beta; PET, positron emission tomography.
[a]The x-axis is reversed into non-pathologic direction (high to low).
[b]The x-axis shows the percentile with respect to proportion in non-pathologic direction.

3.2 | Evaluation of plasma $A\beta_{42}/A\beta_{40}$ as a prescreening tool for a clinical trial

When considering the specific intended use of evaluating plasma $A\beta_{42}/A\beta_{40}$ as a prescreening tool for a secondary prevention trial in asymptomatic $A\beta$+ participants, we assumed the amyloid-positive prevalence will be around 15% when using amyloid PET visual read.[24] Figure 2A shows the risk and the cumulative risk estimates (PPV and 1−NPV) to determine the utility for this intended use. The 1−NPV curve was close to zero for almost 50% of the population, meaning that one could screen out a large proportion of true amyloid-negative individuals without falsely screening out any positives. The PPV, 1−NPV, and percentage screen-out can be directly translated into other trial characteristics, such as total number of patients needing to be screened and number of downstream assessments avoided (e.g., PET/CSF) (Figure 2B). For example, if 45% of the population is screened out, the rate of falsely screened-out amyloid-positives in the screen-out population is 2% (1−NPV), and the amyloid-positive prevalence increases from 15% without prescreening to 25% in the screen-in population (PPV). That translates into needing to prescreen

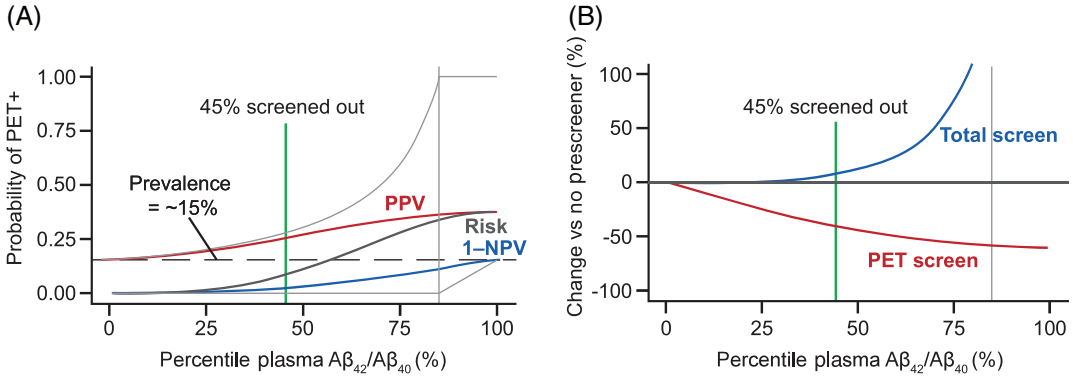

FIGURE 2 Plasma $A\beta_{42}/A\beta_{40}$. (A) Integrated risk-profile plot with an assumed prevalence of 15% showing the estimated probability of being PET+ (risk) in black, PPV in red, and 1-NPV in blue for the percentiles of the plasma $A\beta_{42}/A\beta_{40}$ distribution (reversed).[a] The percentage would correspond to the % screen-out rate. (B) Number needed to screen (total and PET). Aβ, amyloid beta; NPV, negative predictive value; PET, positron emission tomography; PPV, positive predictive value.
[a]The x-axis shows the percentile with respect to proportion in a non-pathologic direction.

7% more participants (to compensate for the falsely screened-out true positives) but reducing the number of PET/CSF assessments by 40%. Figure S4 shows the integrated risk-profile plot of a direct comparison between plasma $A\beta_{42}/A\beta_{40}$, CSF $A\beta_{42}/A\beta_{40}$, and CSF pTau/$A\beta_{42}$.

3.3 | BioFINDER robustness

A robust biomarker is not sensitive to small measurement errors. In the following descriptions, we used cutoffs for a more intuitive illustration of robustness issues; however, the reader should note that the same logic applies without using a cutoff, as there is clinical performance (risk) associated with each biomarker concentration. Figure 3A shows that adding imprecision of 10% CV led to different rates of reclassification for the three biomarker ratios, illustrating how the same level of imprecision can impact the performance of biomarkers differently (26% for plasma $A\beta_{42}/A\beta_{40}$, 4% for CSF $A\beta_{42}/A\beta_{40}$, and 2% for pTau/$A\beta_{42}$).

Figure 3B illustrates that the proportion of data points close to the cutoff is different for plasma $A\beta_{42}/A\beta_{40}$, CSF $A\beta_{42}/A\beta_{40}$, and CSF pTau/$A\beta_{42}$. This was not only driven by better discrimination – while the two CSF ratios had similar clinical discrimination performance (classifying PET+ and PET−), the data points for CSF $A\beta_{42}/A\beta_{40}$ were closer to the cutoff (dotted line) than pTau/$A\beta_{42}$. This is more extreme for plasma $A\beta_{42}/A\beta_{40}$ where the data points are even closer to the cutoff.

Figure 3C on the left compares the range between the three biomarker ratios by dividing each data point by the respective cutoff; 2.0 on the x-axis indicated that a data point was twice as large as the cutoff. For plasma $A\beta_{42}/A\beta_{40}$, the whole distribution had a very narrow range around the cutoff that can already be anticipated by looking at the percentage difference (or fold change) between the means of the two distributions, which is ~ 200% for CSF pTau/$A\beta_{42}$, ~ 100% for CSF $A\beta_{42}/A\beta_{40}$, and ~ 10% for plasma $A\beta_{42}/A\beta_{40}$.

Figure 3C on the right illustrates that a 22% shift (e.g., caused by a 10% shift in opposite directions for the ratio numerator and denominator) moves almost all patients above the cutoff for plasma, meaning all patients are classified as negative and the test loses all clinical performance. In contrast, this had a moderate impact on CSF $A\beta_{42}/A\beta_{40}$ but very little impact on CSF pTau/$A\beta_{42}$, which is why it is considered the biomarker combination with the highest robustness.

For a systematic evaluation, we simulated the impact of imprecision and bias on patient classification. Figure 4 shows how sensitivity and specificity change for varying imprecision (different red and blue scales) and for added bias of the ratio as the result of different biases from the numerator and denominator of the ratio. For plasma $A\beta_{42}/A\beta_{40}$, very small individual bias resulted in a dramatic loss of sensitivity or specificity. For example, with an individual bias of 2% (ratio bias 4%) and no added imprecision, the sensitivity dropped from 83% to 67%; with individual bias of 4% (ratio bias 8.3%), the sensitivity dropped further down to 40%, reaching almost 0% with individual biases of 10% (ratio bias 22%). While the exact amount of allowable bias and imprecision (both components of the TAE) depends on the specific cutoff and intended use, the steep drop in performance indicates that analytical performance requirements (and requirements for other sources of variability, e.g., preanalytical sample handling) would have to be extremely strict for any intended use. As an example, the intended use of prescreening in clinical trials would require an allowable bias of <2% to meet performance acceptance criteria of high sensitivity (otherwise the total number to screen increases drastically). In contrast, the curves for both CSF ratios are much less steep and a moderate increase in imprecision has almost no impact – indicating that it only affects patients in an area around the cutoff where there is naturally a more mixed population of positives and negatives; reclassification of those does not change the clinical performance.

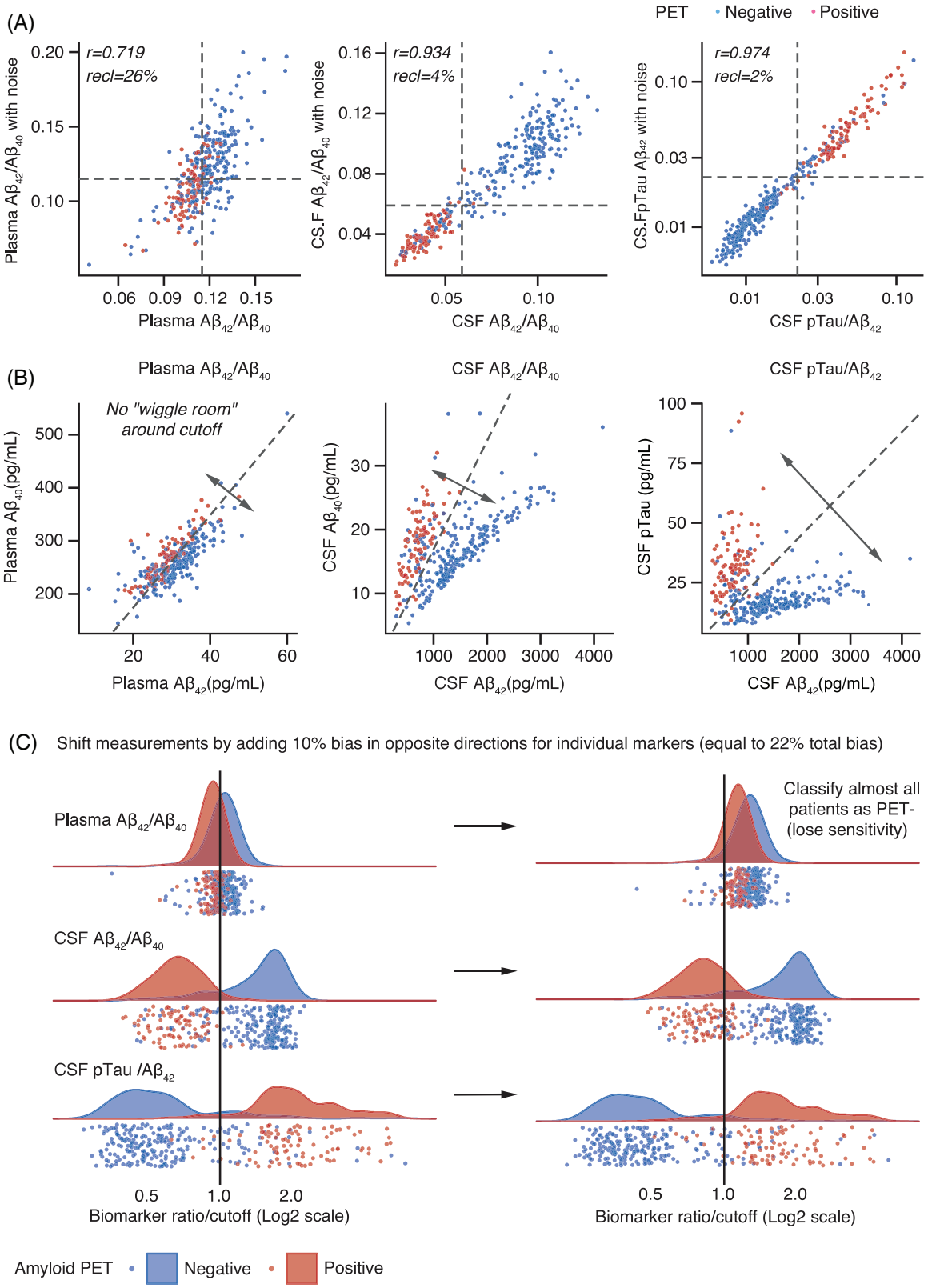

FIGURE 3 (A) Illustration of CV: scatterplots of ratio biomarkers with and without 10% added noise (CV). (B) Scatterplots of individual biomarkers that form a ratio to illustrate proximity to cutoff. (C) Densities of ratio biomarkers with and without 10% bias added. Aβ, amyloid beta; CSF, cerebrospinal fluid; CV, coefficient of variation; r, correlation coefficient; PET, positron emission tomography; pTau, phosphorylated tau; recl, reclassification rate at a specific cutoff.

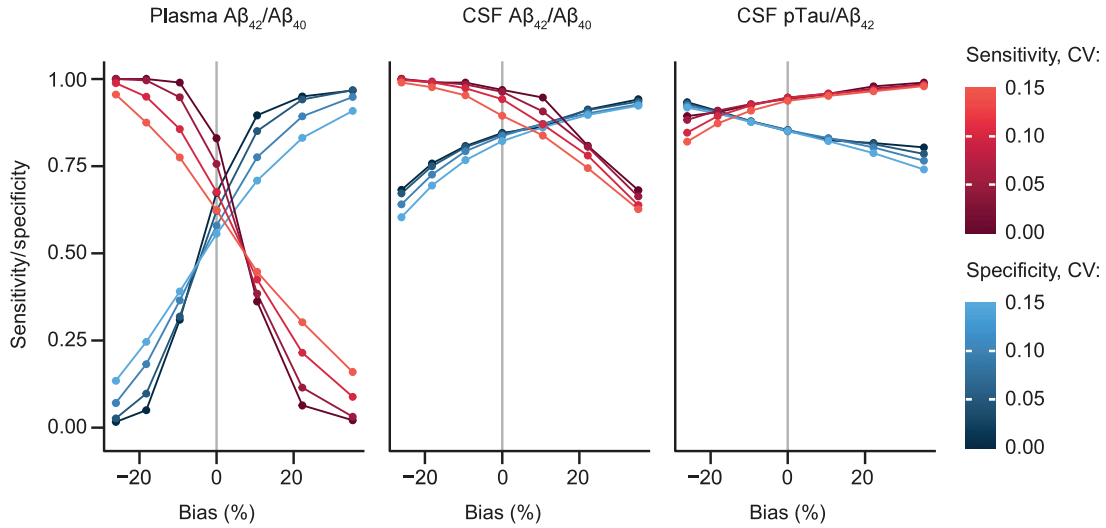

FIGURE 4 Impact on sensitivity (red scales); specificity (blue scales) with simulated imprecision (color scales); bias (x-axis) for the three biomarker ratios. Aβ, amyloid beta; CSF, cerebrospinal fluid; CV, coefficient of variation; pTau, phosphorylated tau.

3.4 | Platform comparison (ADNI FNIH)

The FNIH compared different plasma Aβ_{42}/Aβ_{40} measurement technologies using a small subset of well-characterized samples from the ADNI study. This allowed us to analyze different assays in a head-to-head comparison to evaluate if the narrow range that is causing issues with robustness is specific to a measurement technology (e.g., immunoassay vs mass spectrometry), or if it applies to all methods. While some mass spectrometry methods showed better discrimination performance (Figure 5 [left]), all methods suffered from the same narrow range and low fold change suggesting that similar challenges with robustness, as described for the Elecsys assays, will also apply to other methods (Figure 5 [right]).

4 | DISCUSSION

Our results showed lower concordance between plasma Aβ_{42}/Aβ_{40} and amyloid PET, CSF Aβ_{42}/Aβ_{40}, and CSF pTau/Aβ_{42} than that seen between CSF Aβ_{42}/Aβ_{40} or pTau/Aβ_{42} and PET. This applied to other plasma Aβ_{42}/Aβ_{40} detection technologies as well, even though some mass spectrometry methods show higher AUC than Elecsys.[10] While CSF and PET are often used interchangeably,[20] the lower concordance for plasma Aβ_{42}/Aβ_{40} indicates that the intended use should be carefully evaluated. A marker with lower concordance can still be a promising candidate as a prescreening test, and we showed that Elecsys plasma Aβ_{42}/Aβ_{40} could potentially be used to rule out amyloid pathology in a low prevalence setting. However, because very high-risk estimates are not reached (e.g., not exceeding 40% with a prevalence assumption of 15%), ruling in patients only based on plasma Aβ_{42}/Aβ_{40} does not seem feasible. The risk information might still be useful for triaging patients (e.g., closely monitoring higher-risk individuals).[25,26]

A major limitation of this study is that the evaluation of clinical performance is simplified (i.e., it is performed in a pooled population of CU, SCD, and MCI due to sample size limitations; relies on several assumption; and does not take into account other important factors such as age, Apolipoprotein E4 [ApoE4], and race [available data were predominantly from a Caucasian population]). For example, ApoE4 is a known risk factor for amyloid positivity, and risk estimates based on plasma Aβ_{42}/Aβ_{40} might not be well calibrated without adjusting for ApoE4 status. In clinical routine, decision making rarely relies on one single test or assessment; more extensive evaluations in representative intended-use populations, larger studies, and further assessment of the value of combining plasma Aβ_{42}/Aβ_{40} with other factors are needed. When assessing the robustness of plasma Aβ_{42}/Aβ_{40}, our analysis showed that there is a very high risk of misclassifying patients when only small measurement or pre-analytical errors apply. This is caused by the small difference of plasma Aβ_{42}/Aβ_{40} levels between amyloid-positive and -negative individuals (~ 10%). Therefore, a small measurement error of < 5% can reduce the clinical performance of plasma Aβ_{42}/Aβ_{40} to a level where the biomarker would no longer be clinically useful. For a robust and scalable biomarker (multiple labs, instruments, etc.), small bias and imprecision should not result in drastic changes in the classification of patients; controlling for a small amount of error is challenging in practice, and factors like pre-analytics can only be controlled to a certain extent.

Biases may or may not be correlated between the two markers that form a ratio. In case of Aβ_{42} and Aβ_{40}, biologic and pre-analytical effects are likely to be positively correlated as it often affects both Aβ_{42} and Aβ_{40} but probably not to the same extent.[16] Here, a ratio is helpful, but does not necessarily cancel out all errors. One key pre-analytical factor that has been shown to cause bias with plasma Aβ_{42}/Aβ_{40} is the time to centrifugation of whole blood when stored at room temperature, which causes Aβ_{42}/Aβ_{40} to decrease if not done within 3 to 6 h

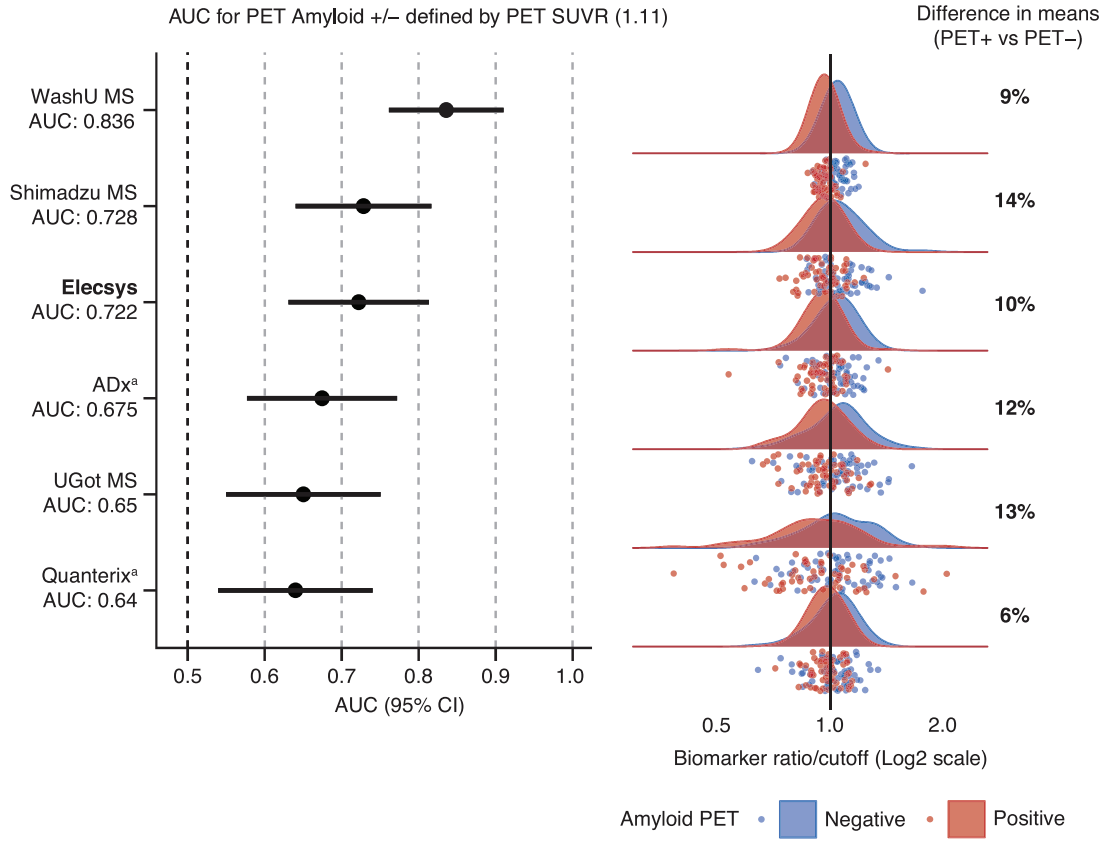

FIGURE 5 ADNI FNIH platform comparison of plasma $A\beta_{42}/A\beta_{40}$ assays. Left panel shows discrimination performance for all markers (AUCs). Right panel shows densities by PET+ and PET–, as well as difference in means (PET+ and PET–). Aβ, amyloid beta; AUC, area under the curve; MS, mass spectrometry; NPV, negative predictive value; PET, positron emission tomography; SUVR, standardized uptake value ratio.
[a]Both ADx and Quanterix used Simoa immunoassays for measurement of Aβ42/Aβ40.

after blood collection.[15,16] This source of error may partially apply in clinical routine where sample handling is not always well standardized compared with a clinical trial. Importantly, bias due to pre-analytical sample handling cannot be controlled by any quality control concept of any measurement technology, since control samples are not affected by it.

We showed that the low dynamic range of plasma $A\beta_{42}/A\beta_{40}$ is not dependent on measurement technology but is inherent in the properties of plasma $A\beta_{42}/A\beta_{40}$ itself. Hence, all plasma $A\beta_{42}/A\beta_{40}$ methods should be carefully assessed for allowable bias and total error for each intended use, to ensure that the risk of misclassifying patients is appropriately controlled for.

A limitation of this study is that the findings have not been derived from a prospective study that incorporates all possible sources of error in the same study where clinical performance is evaluated (e.g., a prospective setting with multiple labs and assay reagent lots, real-time measurements without run-in batches, and pre-analytical conditions as they would apply in clinical routine). All studies so far have been performed retrospectively and without a prespecified cutoff, often normalizing for batch effects,[7–10,27] which minimizes variability compared with the expected variability when using these biomarkers prospectively in clinical trials or routine. Prospective validation under routine conditions is the biggest gap in establishing the clinical utility and scalability (multiple labs, etc.) of plasma $A\beta_{42}/A\beta_{40}$ methods, given the anticipated challenges described.

In conclusion, the narrow range of plasma $A\beta_{42}/A\beta_{40}$ and the small difference between amyloid-positive and -negative subjects results in a very low TAE, which may pose challenges in robustly implementing this biomarker in clinical trials and clinical routine. More research on clinical utility and robustness of blood-based biomarkers in AD in well-designed prospective experiments with prespecified cutoffs is needed. Properties of other promising biomarkers like plasma pTau[28] and glial fibrillary acidic protein[29] should be evaluated as targets for prescreening tools that can be implemented robustly in clinical trials and clinical routine at a large scale.

AUTHOR CONTRIBUTIONS

Christina Rabe, Tobias Bittner, Alexander Jethwa, Ivonne Suridjan, Henrik Zetterberg, Kaj Blennow, and Oskar Hansson designed the experiments. Erik Stomrud and Oskar Hansson collected the samples for BioFINDER1. Christina Rabe, Ekaterina Manuilova, and Michel Friesenhahn analyzed the data. All authors reviewed the data and wrote the manuscript.

ACKNOWLEDGMENTS

H.Z. is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018-02532), the European Research Council (#681712), Swedish State Support for Clinical Research (#ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), the AD Strategic Fund and the Alzheimer's Association (#ADSF-21-831376-C, #ADSF-21-831381-C and #ADSF-21-831377-C), the Olav Thon Foundation, the Erling-Persson Family Foundation, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2019-0228), the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), the European Union Joint Programme for Neurodegenerative Disorders (JPND2021-00694), and the UK Dementia Research Institute at UCL (UKDRI-1003).

O.H. is a Wallenberg Clinical Scholar, and he has received support for the Swedish BioFINDER study from the Swedish Research Council (2016-00906), the Knut and Alice Wallenberg foundation (2017-0383), the Marianne and Marcus Wallenberg foundation (2015.0125), the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation (AF-939932), the Swedish Brain Foundation (FO2021-0293), the Parkinson Foundation of Sweden (1280/20), the Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, the Skåne University Hospital Foundation (2020-O000028), Regionalt Forskningsstöd (2020-0314) and the Swedish federal government under the ALF agreement (2018-Projekt0279).

K.B. is supported by the Swedish Research Council (#2017-00915), the Alzheimer Drug Discovery Foundation (ADDF), USA (#RDAPB-201809-2016615), the Swedish Alzheimer Foundation (#AF-742881), Hjärnfonden, Sweden (#FO2017-0243), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (#ALFGBG-715986), the European Union Joint Programme for Neurodegenerative Disorders (JPND2019-466-236), the National Institutes of Health (NIH), USA, (grant #1R01AG068398-01), and the Alzheimer's Association 2021 Zenith Award (ZEN-21-848495).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Study funded by F. Hoffmann-La Roche Ltd. Medical writing support for the development of this manuscript was furnished by Brittany Gifford, PharmD of Medical Expressions and funded by F. Hoffmann-La Roche Ltd, Basel, Switzerland. COBAS, COBAS E, and ELECSYS are trademarks of Roche.

CONFLICTS OF INTEREST

C.R. and M.F. are full-time employees of Genentech, Inc. (a member of the Roche Group) and owns stock or stock options in Roche.

T.B. is a full-time employee of and owns stock in F. Hoffmann-La Roche Ltd.

A.J. and E.M. are full-time employees of Roche Diagnostics GmbH.

I.S. is a full-time employee and shareholder of Roche Diagnostics International Ltd.

E.S. declares no conflicts of interest.

H.Z. has served at scientific advisory boards and/or as a consultant for AbbVie, Alector, Annexon, Artery Therapeutics, AZTherapies, CogRx, Denali, Eisai, Nervgen, Pinteon Therapeutics, Red Abbey Labs, Passage Bio, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, Biogen, and Roche, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program.

K.B. has served as a consultant and at scientific advisory boards and/or data monitoring committees for Abcam, Axon, Biogen, JOMDD/Shimadzu, Julius Clinical, Lilly, MagQu, Novartis, Prothena, Roche Diagnostics, and Siemens Healthineers, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program.

O.H. has acquired research support (for the institution) from ADx, AVID Radiopharmaceuticals, Biogen, Eli Lilly, Eisai, Fujirebio, GE Healthcare, Pfizer, and Roche. In the past 2 years, he has received consultancy/speaker fees from Amylyx, Alzpath, Biogen, Cerveau, Fujirebio, Genentech, Novartis, Roche, and Siemens. Author disclosures are available in the supporting information.

REFERENCES

1. Alzheimer's Association. 2021 Alzheimer's disease facts and figures. *Alzheimers Dement* 2021;17:327-406.
2. Scheltens P, Blennow K, Breteler MMB, et al. Alzheimer's disease. *Lancet* 2016;388:505-17.
3. Hansson O. Biomarkers for neurodegenerative diseases. *Nat Med* 2021;27:954-63.
4. Blennow K. A review of fluid biomarkers for Alzheimer's disease: moving from CSF to blood. *Neurol Ther* 2017;6:15-24.

5. Mattke S, Hanson M. Expected wait times for access to a disease-modifying Alzheimer's treatment in the United States. *Alzheimers Dement* 2021;epub ahead of print.
6. Mattke S, Cho SK, Bittner T, Hlávka J, Hanson M. Blood-based biomarkers for Alzheimer's pathology and the diagnostic process for a disease-modifying treatment: Projecting the impact on the cost and wait times. *Alzheimers Dement (Amst)* 2020;12:e12081.
7. Nakamura A, Kaneko N, Villemagne VL, et al. High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature* 2018;554:249-54.
8. Ovod V, Ramsey KN, Mawuenyega KG, et al. Amyloid β concentrations and stable isotope labeling kinetics of human plasma specific to central nervous system amyloidosis. *Alzheimers Dement* 2017;13:841-9.
9. Palmqvist S, Janelidze S, Stomrud E, et al. Performance of fully automated plasma assays as screening tests for alzheimer disease-related β-amyloid status. *JAMA Neurol* 2019;76:1060-9.
10. Janelidze S, Teunissen CE, Zetterberg H, et al. Head-to-head comparison of 8 plasma amyloid-β 42/40 assays in Alzheimer disease. *JAMA Neurol* 2021;78:1375-82.
11. Pepe M. *The Statistical Evaluation of Medical Tests for Classification and Prediction*: Oxford University Press; 2004.
12. Pepe MS, Feng Z, Huang Y, et al. Integrating the predictiveness of a marker with its performance as a classifier. *Am J Epidemiol* 2007;167:362-8.
13. Smith AF, Shinkins B, Hall PS, Hulme CT, Messenger MP. Toward a framework for outcome-based analytical performance specifications: A methodology review of indirect methods for evaluating the impact of measurement uncertainty on clinical outcomes. *Clin Chem* 2019;65:1363-74.
14. Benedet AL, Brum WS, Hansson O, et al. The accuracy and robustness of plasma biomarker models for amyloid PET positivity. *Alzheimers Res Ther* 2022;14:26.
15. Rózga M, Bittner T, Batrla R, Karl J. Preanalytical sample handling recommendations for Alzheimer's disease plasma biomarkers. *Alzheimers Dement (Amst)* 2019;11:291-300.
16. Verberk IMW, Misdorp EO, Koelewijn J, et al. Characterization of pre-analytical sample handling effects on a panel of Alzheimer's disease–related blood-based biomarkers: Results from the Standardization of Alzheimer's Blood Biomarkers (SABB) working group. Alzheimers Dement;epub ahead of print.
17. Petersen PH, Jensen EA, Brandslund I. Analytical performance, reference values and decision limits. A need to differentiate between reference intervals and decision limits and to define analytical quality specifications. *Clin Chem Lab Med* 2011;50:819-31.
18. Petrazzuoli F, Vestberg S, Midlöv P, Thulesius H, Stomrud E, Palmqvist S. Brief cognitive tests used in primary care cannot accurately differentiate mild cognitive impairment from subjective cognitive decline. *J Alzheimers Dis* 2020;75:1191-201.
19. Foundation for the National Institutes of Health (FNIH). Biomarkers consortium - plasma Aβ as a predictor of amyloid positivity in Alzheimer's Disease. Accessed April 4, 2022. https://fnih.org/our-programs/biomarkers-consortium/programs/plasma-abeta
20. Hansson O, Seibyl J, Stomrud E, et al. CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement* 2018;14:1470-81.
21. Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI FNIH method sheet. Accessed April 4, 2022. http://adni.loni.usc.edu/data-samples/access-data/
22. Frank Harrell. Clinicians' misunderstanding of probabilities makes them like backwards probabilities such as sensitivity, specificity, and type I error. Accessed April 4, 2022. https://www.fharrell.com/post/backwards-probs/
23. Westgard JO. Useful measures and models for analytical quality management in medical laboratories. *Clin Chem Lab Med* 2016;54:223-33.
24. Sperling RA, Donohue MC, Raman R, et al. Association of factors with elevated amyloid burden in clinically normal older individuals. *JAMA Neurol* 2020;77:735-45.
25. O'Bryant SE, Mielke MM, Rissman RA, et al. Blood-based biomarkers in Alzheimer disease: Current state of the science and a novel collaborative paradigm for advancing from discovery to clinic. *Alzheimers Dement* 2017;13:45-58.
26. Hampel H, O'Bryant SE, Molinuevo JL, et al. Blood-based biomarkers for Alzheimer disease: Mapping the road to the clinic. *Nat Rev Neurol* 2018;14:639-52.
27. West T, Kirmess KM, Meyer MR, et al. A blood-based diagnostic test incorporating plasma Aβ42/40 ratio, ApoE proteotype, and age accurately identifies brain amyloid status: Findings from a multi cohort validity analysis. *Mol Neurodegener* 2021;16:30.
28. Palmqvist S, Janelidze S, Quiroz YT, et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *JAMA* 2020;324:772-81.
29. Chatterjee P, Pedrini S, Stoops E, et al. Plasma glial fibrillary acidic protein is elevated in cognitively normal older adults at risk of Alzheimer's disease. *Transl Psychiatry* 2021;11:27.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rabe C, Bittner T, Jethwa A, et al. Clinical performance and robustness evaluation of plasma amyloid-$\beta_{42/40}$ prescreening. *Alzheimer's Dement*. 2023;19:1393–1402. https://doi.org/10.1002/alz.12801