

A two-step workflow based on plasma p-tau217 to screen for amyloid β positivity with further confirmatory testing only in uncertain cases

Received: 11 January 2023

Accepted: 18 July 2023

Published online: 31 August 2023

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Cost-effective strategies for identifying amyloid- β (A β) positivity in patients with cognitive impairment are urgently needed with recent approvals of anti-A β immunotherapies for Alzheimer's disease (AD). Blood biomarkers can accurately detect AD pathology, but it is unclear whether their incorporation into a full diagnostic workflow can reduce the number of confirmatory cerebrospinal fluid (CSF) or positron emission tomography (PET) tests needed while accurately classifying patients. We evaluated a two-step workflow for determining A β -PET status in patients with mild cognitive impairment (MCI) from two independent memory clinic-based cohorts ($n = 348$). A blood-based model including plasma tau protein 217 (p-tau217), age and *APOE* $\epsilon 4$ status was developed in BioFINDER-1 (area under the curve (AUC) = 89.3%) and validated in BioFINDER-2 (AUC = 94.3%). In step 1, the blood-based model was used to stratify the patients into low, intermediate or high risk of A β -PET positivity. In step 2, we assumed referral only of intermediate-risk patients to CSF A β 42/A β 40 testing, whereas step 1 alone determined A β -status for low- and high-risk groups. Depending on whether lenient, moderate or stringent thresholds were used in step 1, the two-step workflow overall accuracy for detecting A β -PET status was 88.2%, 90.5% and 92.0%, respectively, while reducing the number of necessary CSF tests by 85.9%, 72.7% and 61.2%, respectively. In secondary analyses, an adapted version of the BioFINDER-1 model led to successful validation of the two-step workflow with a different plasma p-tau217 immunoassay in patients with cognitive impairment from the TRIAD cohort ($n = 84$). In conclusion, using a plasma p-tau217-based model for risk stratification of patients with MCI can substantially reduce the need for confirmatory testing while accurately classifying patients, offering a cost-effective strategy to detect AD in memory clinic settings.

AD is the primary cause of dementia and is neuropathologically defined by the accumulation of extracellular A β plaques and intracellular tangles of hyperphosphorylated tau¹⁻³. Established AD biomarkers are essential for patient management and will become increasingly

important as disease-modifying treatments approach clinical practice⁴. New anti-A β therapies have shown promising results in clearing A β from the brain⁵⁻⁷, leading to approvals of aducanumab and lecanemab by the US Food and Drug Administration (FDA). Confirmation of underlying

✉ e-mail: wagner.scheeren.brum@gu.se; oskar.hansson@med.lu.se

AD biomarker abnormalities will be key in determining eligibility for disease-modifying treatments in patients with cognitive impairment visiting memory clinics⁸. Nevertheless, the high cost, invasiveness, time-consuming nature and limited availability of CSF and PET biomarkers hamper their widespread use to screen for AD biomarker positivity in memory clinics.

Blood-based biomarkers hold promise to aid in delivering a biomarker-supported AD diagnosis in a minimally invasive and scalable manner⁴. Plasma p-tau species, including p-tau181, p-tau217 and p-tau231, have shown high performance to identify underlying AD^{9–11}. Plasma p-tau217 (tau phosphorylated at Thr217) shows the highest fold-changes in Aβ-positive patients with cognitive impairment, thus being less susceptible to analytical variation^{10,12–14}. Moreover, plasma p-tau217 is strongly associated with measures of Aβ pathology and its levels change before tau-PET abnormalities are detectable in AD progression^{15–17}, making it a feasible candidate to implement as a routine clinical chemistry test to screen for Aβ positivity in memory clinics.

Nevertheless, the implementation of new AD blood biomarkers into a comprehensive diagnostic workflow for detecting Aβ positivity has received less attention, and the Alzheimer's Association guidelines for appropriate use of AD blood biomarkers recently highlighted the need for objectively evaluating such a strategy¹⁸. Indeed, even the best-performing blood p-tau biomarkers present a higher group-level overlap than established CSF and PET biomarkers^{19,20}. Consequently, handling their results more granularly could potentially reduce the burden of submitting most patients to confirmatory CSF or PET testing. In this context, a model-based approach for interpreting biomarkers alongside clinically relevant information, which is a common strategy in several medical areas^{21,22}, might also be well suited when screening for AD^{23–25}.

In two independent secondary memory clinic-based cohorts, we evaluated a two-step workflow for detecting brain amyloidosis (as indexed by Aβ-PET) in patients with MCI. Step 1 consisted of a diagnostic model based on plasma p-tau217, age and *APOE* ε4 (apolipoprotein E allele ε4) for risk stratification of Aβ-PET positivity. Step 2 was based on confirmatory testing with CSF Aβ42/Aβ40 only in those patients with uncertain outcomes at step 1. In secondary analyses, this workflow was evaluated using a different plasma p-tau217 immunoassay version in a third cohort, from a distinct geographical setting. We demonstrate that such a two-step workflow can lead to a reduction in the number of confirmatory Aβ tests needed while preserving a high overall accuracy for detecting Aβ-PET status.

Results

Participant characteristics

In total, we included 348 MCI participants from BioFINDER-1 ($n = 136$) and BioFINDER-2 ($n = 212$) (Supplementary Table 1). Frequencies of Aβ-PET positivity (BioFINDER-1, 60.3%; BioFINDER-2, 60.8%) and *APOE* ε4 carriership (BioFINDER-1, 49.3%; BioFINDER-2, 55.2%) were similar and both cohorts had fewer females (BioFINDER-1, 35.3%; BioFINDER-2,

42.0%). Included patients from the two cohorts presented similar Mini-Mental State Examination (MMSE) scores, ages and plasma p-tau217 levels (as measured by the Lilly Research Laboratories' assay unless otherwise specified). Comorbidities were frequent, with frequencies in the combined population ($n = 348$) of 54.0% for cardiovascular disease, 15.8% for diabetes, 37.9% for dyslipidemia and 9.2% for chronic kidney disease (CKD).

Model development, validation and threshold definition

Plasma p-tau217, age and *APOE* ε4 status were evaluated as candidate predictors for developing a logistic regression model for Aβ-PET positivity with bootstrapped backward variable elimination in BioFINDER-1 (Supplementary Table 2). The full model, including plasma p-tau217, age and *APOE* ε4, was selected, presenting an optimism-corrected AUC of 89.3% (95% confidence interval (CI) = 83.7–93.8%) for Aβ-PET positivity in BioFINDER-1. At external validation in BioFINDER-2, an independent cohort, the model also presented high discriminatory performance (AUC = 94.3%, 95% CI = 91.2–97.4%). Next, three different thresholding strategies were explored to classify participants into groups with low, intermediate and high risk of Aβ-PET positivity, based on the plasma p-tau217 model-derived probabilities of Aβ-PET positivity. We defined lower probability thresholds with 90%, 95% and 97.5% sensitivity (to avoid missing detection of patients who are Aβ positive), and higher probability thresholds with 90%, 95% and 97.5% specificity (to avoid classifying patients who are Aβ negative as 'high risk'). As the model validated well and displayed good calibration, probability thresholds were derived for the combined BioFINDER-1 and BioFINDER-2 dataset ($n = 348$) (Extended Data Fig. 1). Predicted probabilities of Aβ-PET positivity and the resulting thresholds are shown in Fig. 1a.

Step 1: risk stratification with the blood-based model

Next, we assessed the performance of such thresholds according to the Aβ-PET status (Table 1). We evaluated the accuracy of the low-risk thresholds (specifically, the negative predictive values) by determining the percentage of individuals who are Aβ-PET negative falling below the three different low-risk thresholds. For the more lenient (sensitivity (Se), 90%), intermediate-stringency (Se, 95%) and most stringent (Se, 97.5%) low-risk thresholds evaluated, the accuracy for Aβ-PET negativity was, respectively, 82.0% (18% false negatives), 89.0% (11.0% false negatives) and 93.4% (6.6% false negatives). The accuracy for Aβ-PET positivity (the positive predictive values) of the high-risk thresholds was evaluated by determining the percentage of individuals who are Aβ-PET positive above the different high-risk thresholds. For the more lenient (specificity (Sp), 90%), intermediate-stringency (Sp, 95%) and most stringent (Sp, 97.5%) high-risk thresholds evaluated, the accuracy for Aβ-PET positivity was, respectively, 92.2% (7.8% false positives), 95.2% (4.8% false positives) and 97.7% (2.3% false positives).

When performing risk stratification, the same sensitivity and specificity thresholds were always tested together (for example, 90% Se with 90% Sp, referred to as Se/Sp 90%). As expected, the size of the

Fig. 1 | Development and validation of a two-step workflow for Aβ-PET status capable of reducing further confirmatory tests while accurately classifying patients. **a**, Distribution of predicted probabilities of Aβ-PET positivity based on a logistic regression model including plasma p-tau217, age and *APOE* ε4 status as predictors. The predicted probabilities are displayed for the BioFINDER-1 (model training; left), BioFINDER-2 (model validation; middle) and both combined cohorts (right), with blue dots corresponding to individuals who are Aβ-PET negative and red dots to individuals who are Aβ-PET positive. On the right y axis, the probability values corresponding to the evaluated risk thresholds are demonstrated and accompanied by the metric used to define them (90%, 95%, 97.5% sensitivity or 90%, 95%, 97.5% specificity). The lower dashed line demonstrates where the 95% sensitivity low-risk threshold falls on the probability distribution, with the upper line corresponding to the 95% specificity high-risk threshold. **b**, Flowchart recapitulating results from the first step of the workflow

(blood-biomarker-based risk stratification) and demonstrating the accuracy for the second step of the clinical workflow, when intermediate-risk individuals are referred to lumbar puncture (LP) to perform a CSF Aβ42/Aβ40 test for predicting Aβ-PET status based on the 95% Se/Sp strategy, with the flowchart for the two other strategies presented in Supplementary Information. **c**, The accuracy of both workflow steps combined, corresponding to the proportion of correct classifications for the low- and high-risk groups, along with the proportion of correct CSF Aβ42/Aβ40 classifications in the intermediate-risk group, according to each of the strategies, computed in the BioFINDER-1 and BioFINDER-2 MCI combined populations ($n = 348$). The error bars correspond to 95% CIs. **d**, Dots and lines indicating the observed percentage of reduction in further tests (here using CSF Aβ42/Aβ40) by applying the blood-based risk stratification strategy, based on each of the risk threshold strategies (90% Se/Sp, $n = 301$; 95% Se/Sp, $n = 247$; 97.5% Se/Sp, $n = 205$).

intermediate-risk group increased when more stringent screening strategies were used: with the more lenient strategy of paired Se/Sp 90% thresholds, 13.5% ($n = 47$ out of 348) of individuals were classified as intermediate risk using the blood-based model; with the Se/Sp 95% thresholding strategy, 29.0% of individuals with MCI ($n = 101$ out of 348) fell into the intermediate-risk group; and with the most stringent

strategy of Se/Sp 97.5% thresholds, a larger proportion of individuals, 41.1% ($n = 143$ out of 348), was classified as intermediate risk. For each strategy, the summed percentage of individuals classified into the low- or high-risk groups corresponds to the proportion of patients not needing a confirmatory CSF test, discussed in detail below, alongside overall workflow accuracy.

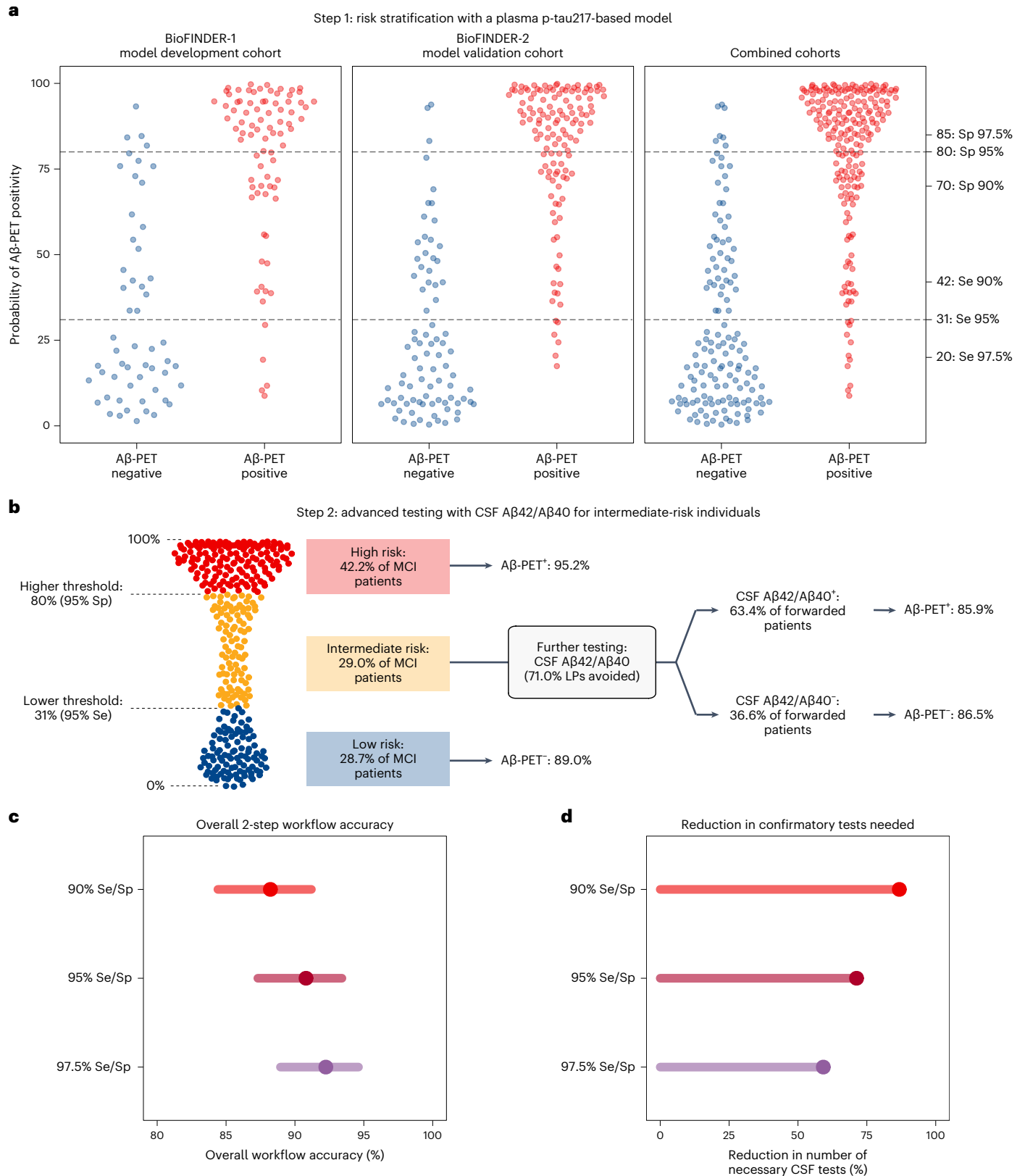


Table 1 | Model-based risk stratification for A β -PET positivity according to the three threshold strategies evaluated

Risk groups	Participants in each risk group (n)	Within-risk group A β -PET status	
		A β -PET negative (n (%))	A β -PET positive (n (%))
90% Se lower-risk threshold (%) 90% Sp higher-risk threshold (%)			
Low risk (<42)	122	100 (82.0)	22 (18.0)
Intermediate risk (42–70)	47	23 (48.9)	24 (51.1)
High risk (>70)	179	14 (7.8)	165 (92.2)
95% Se lower-risk threshold (%) 95% Sp higher-risk threshold (%)			
Low risk (<31)	100	89 (89.0)	11 (11.0)
Intermediate risk (31–80)	101	41 (40.6)	60 (59.4)
High risk (>80)	147	7 (4.8)	140 (95.2)
97.5% Se lower-risk threshold (%) 97.5% Sp higher-risk threshold (%)			
Low risk (<20)	76	71 (93.4)	5 (6.6)
Intermediate risk (20–85)	143	63 (44.1)	80 (55.9)
High risk (>85)	129	3 (2.3)	126 (97.7)

Data are presented as *n* or *n* (%). The first column indicates each of the evaluated strategies for blood-biomarker-based risk stratification, with each strategy's probability threshold indicated in parenthesis next to the low-, intermediate- and high-risk groups. For each of the thresholding strategies, the second column corresponds to the number of screened individuals falling in each risk category. Lastly, A β status is shown for low-, intermediate- and high-risk groups. The percentage of A β negatives in the low-risk group and the percentage of A β positives in the high-risk group correspond to each evaluated threshold's NPV and PPV, respectively.

Step 2: effect of CSF tests for the intermediate-risk group

Considering that the patients classified as intermediate risk at the blood-based risk stratification step were patients with uncertain blood-biomarker outcomes, where A β -PET positivity ranged from 51% to 59%, we investigated whether fully automated CSF A β 42/A β 40 tests would accurately determine the A β -PET status in this subgroup. This approach led to a high concordance between a CSF A β 42/A β 40 and A β -PET status in this group of patients. For the 13.5% of patients with MCI in the intermediate-risk group when using the Se/Sp 90% strategy of the blood-based model, a positive CSF A β 42/A β 40 test had a positive predictive value (PPV) of 82.8% for A β -PET positivity, whereas a negative CSF A β 42/A β 40 test had a negative predictive value (NPV) of 100.0% for A β -PET negativity (Extended Data Fig. 2a). For the Se/Sp 95% blood-based stratification strategy, 29.0% of patients with MCI fell into the intermediate-risk group and CSF A β 42/A β 40 showed a PPV of 85.9% for A β -PET positivity and an 86.5 NPV for A β -PET negativity (Fig. 1b). For the 41.1% of patients with MCI classified as intermediate risk with the Se/Sp 97.5% strategy, CSF A β 42/A β 40 showed a PPV of 87.7% for A β -PET positivity and an 85.5% NPV for A β -PET negativity (Extended Data Fig. 2b). In a sensitivity analysis comparing alternative CSF biomarkers to determine A β -PET status in this uncertain group, A β 42/A β 40 remained as the biomarker with the highest overall accuracy compared with A β 42 alone or p-tau181/A β 42 (Supplementary Table 3).

Workflow overall accuracy and reduction in necessary CSF tests

In general, more stringent screening strategies led to a higher workflow accuracy (Fig. 1c), but also increased the size of the intermediate-risk group who needed further testing (Fig. 1d). When applying the more lenient screening strategy (Se/Sp 90%), the total proportion of correct A β -PET status classifications achieved by the whole two-step workflow (that is, correct blood-based classifications for low- and high-risk groups plus correct CSF A β 42/A β 40 classifications for the intermediate-risk group) was 88.2% (95% CI = 84.4–91.2%). Furthermore, this approach reduced the number of patients needed to be referred for a lumbar puncture by 85.9%. With the Se/Sp 95% risk stratification

strategy, the overall accuracy of the two-step workflow increased to 90.5% (95% CI = 87.3–93.4%), while reducing the number of patients who needed confirmatory CSF testing by 72.7%. The more stringent screening strategy (Se/Sp 97.5%) presented the highest overall workflow accuracy of 92.7% (95% CI = 88.9–94.6%), while still reducing the number of patients who needed to be referred to confirmatory testing by 61.2%. Accuracies for each of the workflow steps are presented separately in Extended Data Fig. 3.

Interassay and geographical validation of the workflow

Finally, we re-fitted the original BioFINDER-1 model but replaced plasma p-tau217 concentrations with plasma p-tau217 values z-transformed, based on reference, cognitively unimpaired (CU), A β -negative populations, to enable interassay validation (model details in Supplementary Table 4), with successful interassay and geographical validation (Fig. 2 and Supplementary Tables 5 and 6). In both BioFINDER-1 and BioFINDER-2, use of z-transformed values of plasma p-tau217 showed similar figures to those of the original concentration-based model, with the following results reported for the 95% Se/Sp strategy with the same thresholds from previous analyses. In BioFINDER-1, the workflow based on the z-scored model showed an accuracy of 90.4% (95% CI = 84.3–94.3%) for A β status while reducing further testing by 67.6%. Similarly, when applying this model in BioFINDER-2, the workflow reached an overall accuracy of 91.0% (95% CI = 86.4–94.2%), while reducing the number of necessary confirmatory CSF tests by 71.2%. Furthermore, we used this adapted BioFINDER-1 model to obtain risk probabilities in a sample of patients with cognitive impairment (*n* = 84) from the Translational Biomarkers in Aging and Dementia (TRIAD) cohort (McGill University, Canada) with complete biomarker availability and plasma p-tau217 measured with another immunoassay version (Janssen R&D), z-transformed based on an internal reference sample of CU A β negatives in TRIAD (demographic characteristics in Supplementary Table 7). When applying the model trained in BioFINDER-1 in TRIAD, using the original BioFINDER 95% Se/Sp thresholds, a similarly high overall workflow accuracy was achieved (89.3%, 95% CI = 80.9–94.3%) while reducing the number of necessary confirmatory tests by 67.9%.

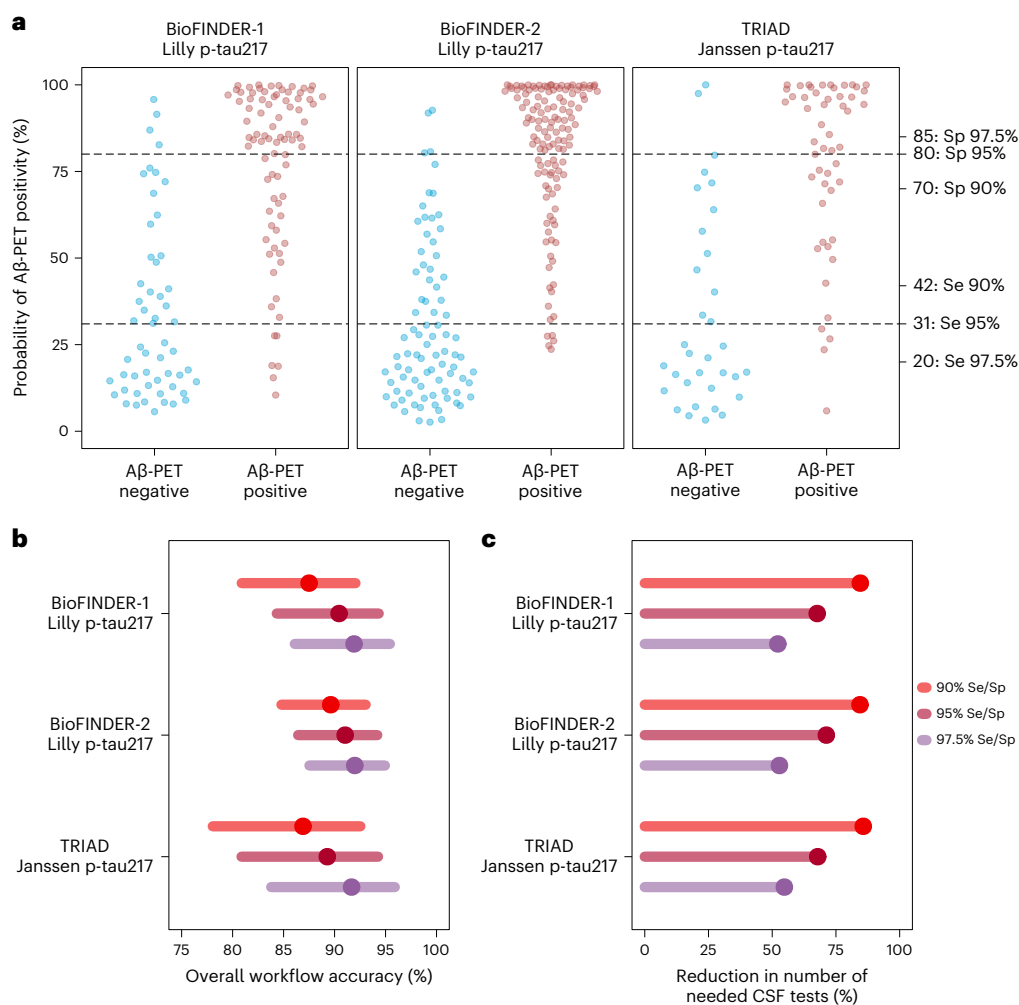


Fig. 2 | A model with z-transformed plasma p-tau217 levels enables interassay and geographical application of the two-step workflow. a, Distribution of predicted probabilities of Aβ-PET positivity based on a logistic regression model including z-transformed plasma p-tau217 levels, combined with age and *APOE* ε4. The z-transformation was done using a CU reference sample from each specific cohort, based on the mean and s.d. of each specific assay in its corresponding population. The predicted probabilities are displayed for the BioFINDER-1 (model training; left), BioFINDER-2 (model validation; middle) and TRIAD (geographical and interassay validation; right), with blue dots corresponding to individuals who are Aβ-PET negative and red dots to individuals who are Aβ-PET positive. On the right y-axis, the probability values corresponding to the evaluated risk thresholds are demonstrated and accompanied by the metric used to define them (90%, 95%, 97.5% sensitivity or 90%, 95%, 97.5% specificity), and the original thresholds from main analyses were used to evaluate their robustness. The lower dashed line demonstrates where the 95% sensitivity low-risk threshold

falls on the probability distribution, with the upper line corresponding to the 95% specificity high-risk threshold. **b**, The accuracy of both workflow steps combined, corresponding to the proportion of correct classifications for the low- and high-risk groups along with the proportion of correct CSF Aβ42/Aβ40 classifications in the intermediate-risk group. The dots correspond to the point estimates for observed accuracy and the lines to 95% CIs, computed based on each cohort's full sample (BioFINDER-1, $n = 136$; BioFINDER-2, $n = 212$; TRIAD, $n = 84$). Each of the threshold strategies is represented by a color as indicated on the right. **c**, The percentage of reduction in further tests by applying the blood-based risk stratification strategy, based on each of the risk threshold strategies. The dots and lines correspond to the observed reduction in needed confirmatory tests (cohort (number of tests avoided according to 90%, 95% and 97.5% strategies, respectively): BioFINDER-1 (115, 92, 71); BioFINDER-2 (179, 151, 112); and TRIAD (72, 57, 46)).

Discussion

In the present study, we evaluated an efficient two-step diagnostic workflow for the identification of brain Aβ-PET status in patients with MCI using risk stratification based on a blood-biomarker model containing plasma p-tau217, age and *APOE* ε4 status (step 1), followed by confirmatory testing with CSF Aβ42/Aβ40 only in patients with intermediate risk at the first blood-based screening step (step 2). In step 1, risk stratification for Aβ-PET positivity was done based on strategies with varying stringency, leading to accurate classifications for Aβ negativity within the low-risk group and for Aβ-positivity in the high-risk group. This was achieved while keeping the intermediate-risk ('uncertain') group reasonably small, substantially reducing the need for further confirmatory testing (reductions from 61.2% to 85.9%).

These results indicate that this workflow might substantially reduce the number of patients who need advanced testing using CSF biomarkers or PET scans, while maintaining a high overall classification accuracy (88.2–92.0%). Furthermore, the two-step workflow showed a similarly high performance when using a different p-tau217 immunoassay in TRIAD, in a different geographical setting. A conceptual flowchart for the future application of the proposed two-step workflow is presented in Fig. 3.

Through this two-step workflow, we propose that one way to implement biomarkers in memory clinics could be by using blood biomarkers in risk-prediction models as a first-line screening tool for patients with memory complaints, provided the clinical presentation warrants an AD-specific blood test. The results achieved with our proposal are in line

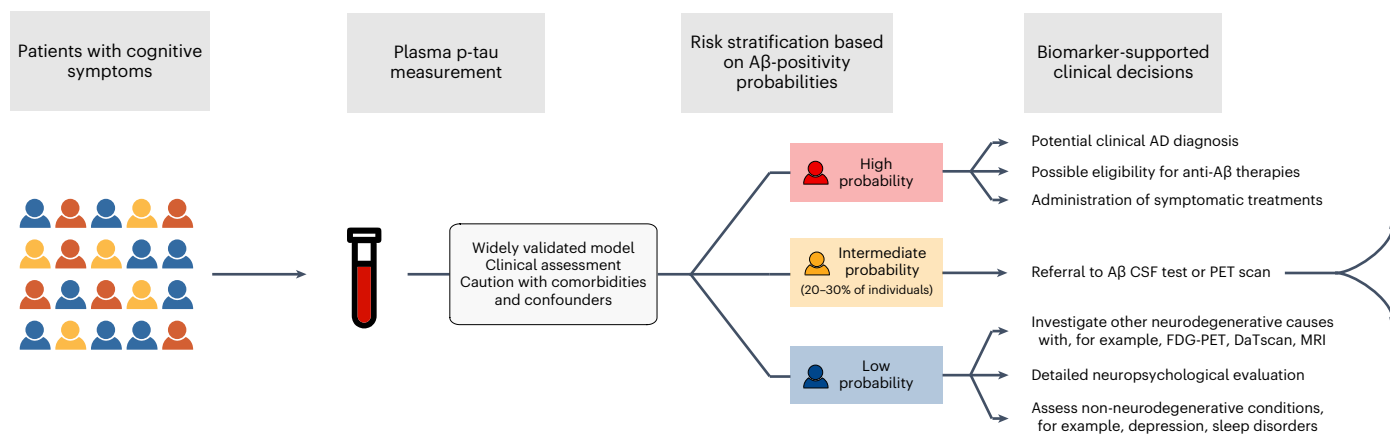


Fig. 3 | A potential workflow for incorporating a plasma p-tau217 risk prediction model for predicting Aβ status in clinical practice. Conceptual flowchart for future implementation of the proposed two-step diagnostic workflow. Participants with cognitive impairment in specialized settings could be screened for the risk of underlying Aβ pathology based on a high-performance plasma p-tau biomarker model also incorporating clinically relevant variables, such as age and *APOE ε4* status. Importantly, a clinical assessment would determine the need for an AD biomarker assessment. Comorbidities potentially affecting circulating biomarker levels should also be taken into consideration.

Based on probability thresholds, chosen according to the decision to be made by the physician, patients could be stratified into low, intermediate and high risk of harboring underlying cerebral Aβ pathology. This biomarker-supported risk stratification could enable highly accurate decisions for individuals in the low- and high-risk groups. Individuals falling within the intermediate-risk group should be forwarded for further testing to determine their Aβ status with a confirmatory PET or CSF Aβ test, depending on center preference and availability. Such a strategy would largely reduce the number of further tests needed, while maintaining a high classification accuracy.

with the recent Alzheimer's Association guidelines on the appropriate use of blood biomarkers, which stated that one of the challenges of the field was to evaluate whether blood-based assessments for AD pathology could achieve high accuracy (90–95%) so that only uncertain cases would be referred for confirmatory CSF or PET tests¹⁸. Although the blood-based model and thresholds herein evaluated are not intended to be the final ones to be used in clinical practice, the evaluated strategies provide a practical example that more rigorous screening thresholds lead to higher accuracy, but simultaneously require advanced testing to be done in more patients. Considering the high accuracy both to rule in and to rule out AD observed in step 1 with these example threshold strategies, we assumed that blood-biomarker-supported decisions could be made for participants in the low- and high-risk groups.

Clinical decisions for the participants within the low-risk group could vary. Depending on clinical manifestations, patients could come back to the memory clinic in 6–24 months for another assessment and blood draw. Alternatively, patients and caregivers could be reassured that AD is unlikely to be the cause of the symptoms and investigation of whether the patient has another neurodegenerative disease would be warranted. For instance, an [¹⁸F]fluorodeoxyglucose (FDG) PET scan could be appropriate for patients with a suspected frontotemporal dementia disorder, a dopamine transporter scan (DaTscan) for those with possible Lewy body etiology and magnetic resonance imaging (MRI) for patients with suspected vascular dementia. In cases where a non-neurodegenerative cause is suspected, a detailed investigation could include further neuropsychological testing and should focus on other possible (and sometimes reversible) causes of worsening in cognitive function, such as depression, post-traumatic stress disorder, substance abuse, delirium, sleep apnea and so on (Fig. 3).

High-risk participants who are very likely to have AD because the etiology causing the symptoms could be clinically diagnosed with greater confidence, allowing for quicker initiation of available treatments than if CSF or PET testing was required. This applies to current symptomatic treatments and, potentially, to new disease-modifying therapies. Even when anti-Aβ therapies obtain coverage by health systems globally, Aβ-PET might not always be a clinical option given the high costs and limited availability. Thus, determining the feasibility of delivering new therapies solely based on blood biomarkers and related algorithms is needed. Ongoing trials, such as TRAILBLAZER-3

(NCT05026866), enrolling participants only with plasma p-tau217, will further aid in clarifying whether anti-Aβ immunotherapies can potentially be delivered without advanced testing. It is important to note that using plasma p-tau217 in a screening diagnostic model alongside other predictors does not exclude the need for interpreting biomarker concentration results alone, because they closely reflect dynamic brain pathological changes, and evaluation of their concentrations alone could also be useful to clinically monitor disease progression and treatment response in the future^{26,27}.

In the second step of the workflow, we evaluated CSF Aβ42/Aβ40 as a confirmatory diagnostic test of Aβ-PET status in patients with uncertain (intermediate-risk), blood-biomarker-based outcomes. On widespread implementation of such a workflow, the choice confirmatory test will depend on patient and physician preferences, as well as center availability. CSF testing has the advantage of being simpler and more widely available in secondary memory clinics due to its low infrastructural complexity, in comparison to imaging procedures. In centers where lumbar punctures are not usually performed and a PET scan is not a possible referral, patients could be referred to a tertiary clinic for a lumbar puncture. Costs for Aβ-PET might still be a complicating factor, because it is still mostly used in research and healthcare system coverage is still limited for clinical purposes, as in the USA²⁸, whereas CSF tests are covered and widely used in European countries, for instance²⁹.

Plasma p-tau217 was chosen as the main blood-biomarker predictor in the screening model for Aβ positivity for being a robust AD-specific biomarker with a large fold-change in Aβ positive patients with cognitive impairment¹⁰, consistently outperforming other p-tau markers in comparison studies^{12,13,30}. As tangle accumulation is more associated with cognitive worsening in the symptomatic phases of AD, another advantage of p-tau217 is that it seems to be driven by both Aβ and tau pathologies³¹. Other blood biomarkers such as p-tau231 and Aβ42/Aβ40 seem to plateau with early Aβ accumulation, besides potential robustness issues due to the very limited AD-related fold-change (around an 8–14% reduction) for the latter^{32,33}, compared with fold-changes usually >200% for different plasma p-tau217 assays^{10,12,13}. Although it is not yet determined which plasma p-tau217 assays will be implemented on a large scale, we demonstrated the workflow's performance to be robust using two different, validated, p-tau217

immunoassays^{12,13,34}. This shows that such a model could potentially be used based on the locally available plasma p-tau217 assay, with biomarker levels z-transformed based on each center's cognitively unimpaired A β -negative reference sample. Both immunoassays demonstrated comparable performance across cohorts (with wider CIs in TRIAD due to lower sample size), although specific assay comparisons were not within the scope of the present work. Importantly, the probability thresholds derived in the concentration-based model performed well between assays without the need for re-optimization, with the workflow demonstrating similar performance both within two independent cohorts from the same geographical setting (BioFINDER-1 and BioFINDER-2) and in a memory clinic-based cohort from a different continent (TRIAD).

Previous reports indicate that, although with varying effect sizes, CKD might be positively associated with plasma p-tau levels^{35–37}. Indeed, we found a higher frequency of CKD in the false-positive group with the 95% Se/Sp strategy with the plasma p-tau217 (Lilly) original model (Supplementary Tables 8 and 9). However, misclassifications were not frequent and generally occurred throughout the whole span of renal function, with most of the misclassified patients with CKD being, in fact, close to the estimated glomerular filtration rate cutoff for abnormal renal function (Extended Data Fig. 4). Furthermore, these false-positive patients with CKD often showed up as CSF positive for A β 42/A β 40 with elevated CSF p-tau levels (Extended Data Fig. 5), possibly suggesting an early disease process rather than a peripheral confounding effect. Although these and previous results may nevertheless recommend some caution when interpreting plasma p-tau in patients with comorbidities, it seems difficult to determine whether reduced renal function might have truly impacted false positivity in our study in light of the above-mentioned patient-level information.

Traditionally, CSF and PET diagnostic biomarkers for AD have been interpreted by clinicians as binary results (normal/abnormal) and they have not been largely used for risk stratification with prediction models. Despite being excellent proxies of AD pathology, new p-tau blood biomarkers do not present a clear bimodal distribution between non-AD and AD groups and, importantly, they present higher group-level overlaps than CSF and PET A β biomarkers^{10,38,19}. In consequence, searching for an 'optimal' binary cutoff for blood biomarkers might be difficult. In this context, inclusion of other easily accessible variables could help to mitigate the group-overlap issue, and use of different cutoffs with a specific clinical goal (for example, screen-out or screen-in AD) might improve their clinical use³⁹. In our and previous studies evaluating blood-biomarker models^{23–25}, including age and *APOE* ϵ 4 status—known relevant risk factors of A β positivity^{40,41}—led to more discriminative models with a higher spread in predictions, which can help in supporting better screening decisions, and such models will probably become more common in AD diagnostics. In other medical fields in which risk-prediction models are more frequently used, it is common to combine both condition-related biomarkers with other relevant variables, for example, risk factors and genetic information, such as the HEART score for identifying ischemic etiology of acute chest pain²¹ (combining demographics, risk factors and biomarkers of myocardial damage) and the STHLM3 model for diagnosing prostate cancer²² (combining demographics, genetic polymorphisms and prostate-specific antigen levels).

We acknowledge strengths and limitations of our study. A strength of the present study was that we included a large group of cognitively impaired participants, from three independent memory clinic cohorts from two geographically distinct settings. The workflow showed high performance in patients extensively phenotyped with two different plasma p-tau217 assay variants measured in different analytical platforms, two FDA-approved CSF A β 42/A β 40 assays and two A β -PET radiotracers. Taken together, we consider our design supports the potential generalizability of our findings, although further validation

in diverse populations and settings is warranted. Although we first envision such a workflow to be applied in memory clinics with the capacity to handle advanced testing (CSF and/or PET) and new therapies, this workflow could be most useful in primary care in the future, possibly facilitating the referral process to specialist clinics. We highlight that the BioFINDER-1 and BioFINDER-2 populations in the present study consisted of memory clinic patients referred from primary care, presenting a wide range of comorbidities, and also presenting relatively low educational attainment (median 12 years) and similar age ranges to other aging and memory clinic cohorts^{42–44}. Although the BioFINDER samples had higher proportions of men (and women who are more affected by AD), A β -PET positivity was more frequent in women (65.7%) than in men (57.3%). A limitation of our study is that plasma biomarker measurements, for each of the assays, were conducted in a single-batch manner (as is standard in cohort studies). Before clinical routine implementation, assays and biomarker-based model strategies will have to be prospectively validated. Another limitation is that the ideal reference standard for *in silico* evaluation of such a workflow would have been neuropathology, which is not yet available for the cohorts included, but our reference standard, A β -PET, has been widely validated against neuropathology⁴.

In conclusion, when screening patients with MCI for the presence of A β positivity, performing risk stratification with a plasma p-tau217-based model can lead to highly accurate classifications while substantially reducing the number of patients referred for further costly or invasive A β tests. Implementing such a workflow to detect AD in the future could considerably reduce advanced testing with CSF or PET, minimizing the burden for patients and caregivers, as well as the costs for healthcare providers.

Methods

Participants

In this cross-sectional study, we included patients with MCI from two independent cohorts, based on complete availability of plasma p-tau217, CSF A β 42/A β 40, A β -PET and *APOE* ϵ 4 genotyping. Our model training cohort, BioFINDER-1 (NCT01208675), recruited patients between January 2010 and January 2015 and our validation cohort, BioFINDER-2 (NCT03174938), started recruitment in May 2017. In both cohorts, the patients were consecutively recruited from secondary memory clinics in the southern part of Sweden, where most of the study participants were referred directly from primary care, as described below. In Supplementary Information, we demonstrate that the included BioFINDER-1 and BioFINDER-2 populations (that is, with full biomarker availability) were similar to the nonincluded participants due to lack of data for one or more biomarkers (Supplementary Tables 7 and 8)^{10,45,46}.

The BioFINDER-1 inclusion criteria for enrolling participants with subjective cognitive decline or MCI were as follows: (1) having been referred owing to cognitive symptoms experienced by the participant or perceived by an informant; (2) age between 60 and 80 years; (3) MMSE score of 24–30 points at the baseline visit; (4) do not fulfill the criteria for any dementia; and (5) fluency in Swedish. The exclusion criteria were as follows: (1) a systemic illness or organ failure of substantial severity that would hinder participation in the study; (2) current substance misuse or alcohol abuse; (3) refusal of neuropsychological assessment or lumbar puncture; and (4) cognitive impairment at baseline that could, with high confidence, be explained by another condition or disease, such as major cerebral hemorrhage, normal pressure hydrocephalus, brain tumor, brain infection, epilepsy, multiple sclerosis, psychotic disorders, severe depression or ongoing use of medication that causes a reduction in cognitive functioning (such as high-dose benzodiazepines). The clinical diagnosis was delivered at baseline based on an extensive battery of neuropsychological tests evaluating verbal and episodic memory, visuospatial ability and attention/executive domains, as described in detail elsewhere⁴⁶. In the whole BioFINDER-1 study, for which enrollment was

completed, a thorough analysis on referral origin had been previously conducted as described by Petrazzuoli et al.⁴⁶. Most of the BioFINDER-1 patients (80.8%) were referred from primary care, whereas 12.5% of referrals were made by other specialist clinics and 6.7% of patients were self-referrals⁴⁶. The inclusion criteria for recruitment of patients with MCI for BioFINDER-2 were as follows: (1) aged 40–100 years; (2) referred to the memory clinics due to cognitive symptoms; (3) MMSE score of 24–30 points; (4) did not fulfill the criteria for any dementia (major neurocognitive disorder) according to the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn (DSM-IV)⁴⁷; and (5) fluent in Swedish. The BioFINDER-2 study also recruits patients who are CU, patients with AD dementia and patients with non-AD neurodegenerative conditions, and its general exclusion criteria were as follows: (1) unstable systemic illness that makes it difficult to participate in the study; (2) current alcohol or substance misuse; and (3) refusing lumbar puncture, MRI or PET. Out of the 212 MCI-included participants from BioFINDER-2 with readily available referral data, most were referred from primary care ($n = 179$; 84.4%), followed by hospital referrals ($n = 31$; 14.6%) and self-referrals ($n = 2$; 0.9%).

In both cohorts, a clinical diagnosis of MCI was made for those patients who did not meet the criteria for dementia (major cognitive disorder as in DSM-V⁴⁸) but have lower scores than -1.5 s.d. in at least one cognitive domain such as memory, verbal, attention/executive or visuospatial function. In BioFINDER-1, a senior neuropsychologist made the diagnosis after a thorough neuropsychological battery to make this determination, as previously described⁴⁶. In BioFINDER-2, the MCI diagnosis was based on a score < -1.5 z-scores in any cognitive domain, based on regression normative scores accounting for age, education and test performance in A β -negative controls⁴⁹. The z-scores for each cognitive domain were calculated by averaging the z-scores of relevant tests, with further details on the derivation of such normative equations available elsewhere^{50,51}. The domains included attention/executive function, verbal ability, memory and visuospatial function, and the tests used included Trail Making Test A, Trail Making Test B, Symbol Digit Modalities Test, verbal fluency animals, 15-word short version of the Boston Naming Test, 10-word delayed recall from the Alzheimer's Disease Assessment Scale, and incomplete letters and cube analysis from the Visual Object and Space Perception battery.

In BioFINDER-1 and BioFINDER-2, we also evaluated the presence of comorbidities in the study population, evaluating for history of cardiovascular disease, diabetes or dyslipidemia³⁶. Participants were considered to have cardiovascular disease if they presented with a history of either ischemic heart disease or hypertension, or if they were on anti-hypertensive/cardioprotective therapy. A history of dyslipidemia was considered when patients had such a diagnosis previously made or if they were on lipid-lowering therapy. Participants were considered to have CKD based on estimated glomerular filtration rate < 60 ml min⁻¹ per 1.73 m², accepted as a functional criterion for CKD⁵².

In a secondary analysis, we included a subset of 84 cognitively impaired participants with available plasma p-tau217, CSF A β 42/A β 40, *APOE* ϵ 4 genotype and A β -PET from the TRIAD cohort, recruited from a tertiary care memory clinic specializing in the diagnosis and management of neurodegenerative diseases⁴⁴. All clinical diagnoses were made blinded to biomarker results. All participants had clinical assessments including Clinical Dementia Rating (CDR), MMSE and cerebrovascular disease risk using the Hachinski Ischemic Scale. Participants were excluded from the present study if they had systemic conditions that were not adequately controlled through a stable medication regimen. Other exclusion criteria were active substance abuse, recent head trauma, recent major surgery or MRI/PET safety contraindications. The included participants had MCI as defined based on a CDR of 0.5 and an MMSE between 24 and 30 ($n = 63$), and patients with dementia who had CDR of ≤ 1 ($n = 21$).

All BioFINDER and TRIAD patients gave their written informed consent to participate in the study and participation was voluntary.

The BioFINDER studies were approved by the Ethical Review Board in Lund, Sweden, which is part of the Swedish Ethical Review Authority. TRIAD was approved by the Montreal Neurological Institute PET working committee and the Douglas Mental Health University Institute Research Ethics Board.

Imaging and fluid biomarkers in BioFINDER-1 and BioFINDER-2

A β -PET was quantified using [¹⁸F]flutemetamol on a Philips Gemini TF 16 scanner in BioFINDER-1 and a digital GE Discovery MI scanner in BioFINDER-2. Scans were acquired 90–110 min after the injection of ~ 185 MBq of [¹⁸F]flutemetamol. The standardized uptake value ratio (SUVR) was obtained by normalizing the neocortical composite values to the whole cerebellum as a reference region. FreeSurfer (v.5.3) parcellation of the T1-weighted MR scan was used to transform the PET data to the participants' native T1 space, so as to obtain mean regional SUVR values in predefined neocortical regions of interest, including prefrontal, lateral temporal, parietal, anterior cingulate and posterior cingulate/precuneus⁵³. A β -PET data were binarized into normal and abnormal using cutoffs derived from Gaussian mixture modeling (GMM), with a threshold of ≥ 1.138 for BioFINDER-1 and ≥ 1.033 for BioFINDER-2.

CSF samples were collected and described based on previously described protocols⁵⁴. CSF A β 42/40 was measured using the fully automated Roche Elecsys NeuroTool Kit for the entirety of BioFINDER-1 and for 75% ($n = 161$) of BioFINDER-2 participants^{55,56}. Abnormal CSF status was defined based on previously derived cutoffs determined using GMM, with a threshold of ≤ 0.066 for BioFINDER-1 and ≤ 0.080 for BioFINDER-2 (the higher cutoff in the latter study is due to use of LoBind tubes in BioFINDER-2, according to more recent protocols that prevent A β 42 from binding to the tube walls^{57,58}). For the 25% ($n = 51$) of BioFINDER-2 participants for whom the Elecsys measurements were not available, an abnormal CSF A β 42/40 status was determined using the FDA-approved Lumipulse G assay, with a GMM-derived threshold of ≤ 0.06 (ref. 59). All CSF A β 42/40 measurements were performed at the Clinical Neurochemistry Laboratory, Sahlgrenska Academy.

EDTA plasma samples were collected, handled and processed as previously described^{10,45}. Plasma p-tau217 was quantified using the Mesoscale Discovery platform with an assay developed by Lilly Research Laboratories. Biotinylated-IBA493 was used as a capture antibody and SULFO-TAG-4G10-E2 (anti-tau) as the detector antibody, with sample and antibody dilution at 1:2, as previously described²³. *APOE* ϵ 4 was genotyped using a TaqMan allelic discrimination assay⁶⁰.

Imaging and fluid biomarkers in TRIAD

Individuals were evaluated with plasma p-tau217, CSF A β 42/40 and amyloid-PET using [¹⁸F]JAZD4694. Plasma concentrations of p-tau217 were measured using a Simoa assay developed by Janssen R&D by scientists blinded to clinical, demographic and biomarker information as described previously¹⁶, using the PT3 antibody as capture and HT43 as detector, and samples and detector were diluted 1:2. CSF concentrations of A β 42/40 were quantified using the fully automated Lumipulse G1200 instrument (Fujirebio), with an A β -positivity threshold of 0.068, by scientists blinded to clinical and biomarker information as described previously⁶¹. A [¹⁸F]AZD4694 amyloid-PET-positivity threshold of 1.55 was employed (centiloid ≥ 24), validated based on GMM, CSF thresholds and visual assessments⁶². Blood and CSF collections took place on the same day.

Statistics and reproducibility

First, we developed a logistic regression model using A β -PET status as the outcome with plasma p-tau217, age and *APOE* ϵ 4 status as predictors in BioFINDER-1. Age and *APOE* ϵ 4 were considered as predictors due to their inclusion in recently published, blood-based biomarker models and due to their well-described associations with A β positivity^{23–25,40,41}. Plasma p-tau217 was log-transformed due to its skewed distribution and age was modeled with a linear term. Variables such as

cognitive tests may be of more relevance to prognostic models (that is, predicting cognitive worsening) than in diagnostic models for A β positivity, given the poor association between A β load and symptoms⁶³. To examine whether a simpler model would be preferred to this full model with age, *APOE* ϵ 4 and p-tau217, backward variable deletion was performed during bootstrapped internal validation ($n = 1,000$), with the stopping criterion set at $\alpha = 0.157$, recommended for model development scenarios such as ours⁶⁴. The model most frequently chosen during this procedure was externally validated in BioFINDER-2. For model performance, we used the receiver operating characteristic's AUC. In BioFINDER-1, the optimism-corrected AUC is reported, a metric recommended to account for overfitting-related optimism at model development⁶⁵. Model calibration at external validation was assessed visually⁶⁶. For goodness of fit, we report Nagelkerke's pseudo-coefficient of determination (R^2) and Akaike's information criterion^{65,67}.

Based on the blood biomarker, model-derived probabilities of A β -PET positivity and further testing with CSF A β 42/A β 40, we evaluated a two-step diagnostic workflow. In the first step, different thresholding strategies were explored to classify participants into low-, intermediate- and high-risk groups based on the plasma p-tau217 model-derived probabilities of A β -PET positivity. These strategies were defined based on lower probability thresholds with 90%, 95% and 97.5% sensitivity and higher probability thresholds with 90%, 95% and 97.5% specificity, with the same sensitivity and specificity always being tested together (for example, 90% sensitivity with 90% specificity). For each of the strategies, we calculated the prevalence of A β -PET negativity in the low-risk group along with the prevalence of A β -PET positivity in the high-risk group. For the second step, we tested the scenario in which further testing would be carried out with CSF A β 42/A β 40 measurements only in intermediate-risk participants from the first step. In this group, we reported the concordance between CSF and A β -PET status. Furthermore, we computed the overall workflow accuracy, represented by the proportion of correct A β -PET status classifications in both plasma and CSF steps, as well as the reduction in number of further confirmatory tests by the blood-biomarker-based risk stratification. In a secondary exploratory analysis, we further evaluated the robustness and generalizability of the two-step workflow using z-scored plasma p-tau217 values. The z-scores were obtained based on the distribution of this reference CU A β -negative sample as follows: (plasma p-tau concentration – mean p-tau concentration in CU A β negatives)/(s.d. of plasma p-tau concentration in CU A β -negatives). In BioFINDER-1, z-scored plasma p-tau217 (Lilly) values were obtained based on 283 CU A β -negative older adults with a mean (s.d.) plasma p-tau217 concentration of 0.153 (0.077) pg ml⁻¹. In BioFINDER-2, based on 316 CU A β -negative participants, the mean (s.d.) concentrations were 0.156 (0.064) pg ml⁻¹ for plasma p-tau217 (Lilly). In TRIAD, z-scores were calculated based on 111 A β -negative CU older adults with a mean (s.d.) plasma p-tau217 (Janssen) concentration of 0.052 (0.026) pg ml⁻¹. Such a procedure enables application of the risk-prediction model for different plasma p-tau217 assays, because when z-scored they can be obtained from internal reference samples from clinical chemistry labs and memory clinic services. Briefly, the same original BioFINDER-1 model was re-fitted with z-scored plasma p-tau217 with the Lilly assay. Then, it was validated in two other cohorts: BioFINDER-2, based on z-scored Lilly plasma p-tau217 immunoassay and, in TRIAD, based on z-scored plasma p-tau217 measured with a different p-tau217 immunoassay (Janssen R&D). The whole workflow was re-evaluated for overall accuracy and reduction in the number of advanced tests for all of these secondary analyses, with the same risk thresholds from the original main analysis model. The z-scored model was developed in BioFINDER-1 in the exact same MCI population as that in the main analysis ($n = 136$). When validating the z-scored model in BioFINDER-2 with z-scored Lilly plasma p-tau217, we evaluated it in the exact same BioFINDER-2 MCI population as used in the main analysis ($n = 212$). In TRIAD, the z-scored model was applied in the $n = 84$ patients with

cognitive impairment with key demographic characteristics shown in Supplementary Information. Our sample size was based on complete biomarker availability (for plasma, genetic, CSF and imaging data) rather than on statistically predetermined numbers, but our sample size was similar to those reported in previous publications evaluating risk-prediction models in AD^{23–25}. When applicable, a two-sided α of 0.05 was used and 95% CIs are reported. No data exclusion was performed. Data collection and analysis were not randomized or performed blind to the experimental groups. All statistical analyses were performed in R v.4.1.1 (www.r-project.org), mainly using the 'rms' package⁶⁸.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The present study does not include data available in external or online repositories. Anonymized data will be shared by request from a qualified academic investigator for the sole purpose of replicating procedures and results presented in the article. For BioFINDER, requests will be considered as long as data transfer is in agreement with EU legislation on the general data protection regulation and decisions by the Swedish Ethical Review Authority and Region Skåne, which should be regulated in a material transfer agreement and contact can be made through the study's website (<https://biofinder.se>). Arrangements for data sharing for replication of the findings in the TRIAD dataset are subject to standard data-sharing agreements and further information can be found in the study's website (<https://triad.tnl-mcgill.com>) or via direct contact with study leader pedro.rosa@mcgill.ca.

Code availability

The code that supports the results of the present study is available from the corresponding authors upon request. All models were built using publicly available packages and functions in the R programming language.

References

1. McKhann, G. M. et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **7**, 263–269 (2011).
2. Jack, C. R. et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **14**, 535–562 (2018).
3. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**, e105–e125 (2022).
4. Hansson, O. Biomarkers for neurodegenerative diseases. *Nat. Med.* **27**, 954–963 (2021).
5. van Dyck C. H., et al. Lecanemab in early Alzheimer's disease. *N. Engl. J. Med.* **388**, 9–21 (2023).
6. Mintun, M. A. et al. Donanemab in early Alzheimer's disease. *N. Engl. J. Med.* **384**, 1691–1704 (2021).
7. Budd Haeberlein, S. et al. Two randomized phase 3 studies of aducanumab in early Alzheimer's disease. *J. Prev. Alzheimers Dis.* **9**, 197–210 (2022).
8. Cummings, J. et al. Aducanumab: appropriate use recommendations. *J. Prev. Alzheimers Dis.* **8**, 398–410 (2021).
9. Karikari, T. K. et al. Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: a diagnostic performance and prediction modelling study using data from four prospective cohorts. *Lancet Neurol.* **19**, 422–433 (2020).

10. Palmqvist, S. et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *JAMA* **324**, 772–781 (2020).
11. Ashton, N. J. et al. Plasma p-tau231: a new biomarker for incipient Alzheimer's disease pathology. *Acta Neuropathol.* **141**, 709–724 (2021).
12. Janelidze, S. et al. Head-to-head comparison of 10 plasma phospho-tau assays in prodromal Alzheimer's disease. *Brain J. Neurol.* **146**, 1592–1601 (2022).
13. Ashton, N. J. et al. Plasma and CSF biomarkers in a memory clinic: head-to-head comparison of phosphorylated tau immunoassays. *Alzheimers Dement. J. Alzheimers Assoc.* **19**, 1913–1924 (2023).
14. Cullen, N. C. et al. Test-retest variability of plasma biomarkers in Alzheimer's disease and its effects on clinical prediction models. *Alzheimers Dement. J. Alzheimers Assoc.* **19**, 797–806 (2023).
15. Salvadó, G. et al. Specific associations between plasma biomarkers and postmortem amyloid plaque and tau tangle loads. *EMBO Mol Med.* **15**, e17123 (2023).
16. Therriault, J. et al. Association of phosphorylated tau biomarkers with amyloid positron emission tomography vs tau positron emission tomography. *JAMA Neurol.* **80**, 188–199 (2023).
17. Janelidze, S. et al. Associations of plasma phospho-Tau217 levels With tau positron emission tomography in early Alzheimer disease. *JAMA Neurol.* **78**, 149–156 (2021).
18. Hansson, O. et al. The Alzheimer's Association appropriate use recommendations for blood biomarkers in Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **18**, 2669–2686 (2022).
19. Karikari, T. K. et al. Blood phospho-tau in Alzheimer disease: analysis, interpretation, and clinical utility. *Nat. Rev. Neurol.* **18**, 400–418 (2022).
20. Therriault, J. et al. Equivalence of plasma p-tau217 with cerebrospinal fluid in the diagnosis of Alzheimer's disease. *Alzheimers Dement.* <https://doi.org/10.1002/alz.13026> (2023).
21. Backus, B. E. et al. A prospective validation of the HEART score for chest pain patients at the emergency department. *Int. J. Cardiol.* **168**, 2153–2158 (2013).
22. Grönberg, H. et al. Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. *Lancet Oncol.* **16**, 1667–1676 (2015).
23. Janelidze, S. et al. Detecting amyloid positivity in early Alzheimer's disease using combinations of plasma A β 42/A β 40 and p-tau. *Alzheimers Dement. J. Alzheimers Assoc.* **18**, 283–293 (2022).
24. Hu, Y. et al. Assessment of a plasma amyloid probability score to estimate amyloid positron emission tomography findings among adults with cognitive impairment. *JAMA Netw. Open* **5**, e228392 (2022).
25. Verberk, I. M. W. et al. Plasma amyloid as prescreener for the earliest Alzheimer pathological changes. *Ann. Neurol.* **84**, 648–658 (2018).
26. Pontecorvo, M. J. et al. Association of donanemab treatment with exploratory plasma biomarkers in early symptomatic Alzheimer disease: a secondary analysis of the TRAILBLAZER-ALZ randomized clinical trial. *JAMA Neurol.* **79**, 1250–1259 (2022).
27. Ashton, N. J. et al. Differential roles of A β 42/40, p-tau231 and p-tau217 for Alzheimer's trial selection and disease monitoring. *Nat. Med.* **28**, 2555–2562 (2022).
28. Medicine S of N. SNMMI calls for expanded coverage/ reimbursement of amyloid PET. *J. Nucl. Med.* **63**, 17N (2022).
29. Erickson, P. et al. Prevalence and clinical implications of a β -amyloid-negative, tau-positive cerebrospinal fluid biomarker profile in Alzheimer disease. *JAMA Neurol.* <https://doi.org/10.1001/jamaneurol.2023.2338> (2023).
30. Montoliu-Gaya, L. et al. Mass spectrometric simultaneous quantification of tau species in plasma shows differential associations with amyloid and tau pathologies. *Nat. Aging* **3**, 661–669 (2023).
31. Mattsson-Carlgrén, N. et al. Soluble P-tau217 reflects amyloid and tau pathology and mediates the association of amyloid with tau. *EMBO Mol. Med.* **13**, e14022 (2021).
32. Rabe, C. et al. Clinical performance and robustness evaluation of plasma amyloid- β 42/40 prescreening. *Alzheimers Dement. J. Alzheimers Assoc.* **19**, 1393–1402 (2023).
33. Benedet, A. L. et al. The accuracy and robustness of plasma biomarker models for amyloid PET positivity. *Alzheimers Res. Ther.* **14**, 26 (2022).
34. Groot, C. et al. Diagnostic and prognostic performance to detect Alzheimer's disease and clinical progression of a novel assay for plasma p-tau217. *Alzheimers Res. Ther.* **14**, 67 (2022).
35. Mielke, M. M. et al. Performance of plasma phosphorylated tau 181 and 217 in the community. *Nat. Med.* **28**, 1398–1405 (2022).
36. Pichet Binette, A. et al. Confounding factors of Alzheimer's disease plasma biomarkers and their impact on clinical performance. *Alzheimers Dement.* **19**, 1403–1414 (2023).
37. Janelidze, S., Barthélemy, N. R., He, Y., Bateman, R. J. & Hansson, O. Mitigating the associations of kidney dysfunction with blood biomarkers of Alzheimer disease by using phosphorylated tau to total tau ratios. *JAMA Neurol.* **80**, 516–522 (2023).
38. Leuzy, A. et al. Comparing the clinical utility and diagnostic performance of CSF p-tau181, p-tau217, and p-tau231 assays. *Neurology.* **97**, e1681–e1694 (2021).
39. Wynants, L. et al. Three myths about risk thresholds for prediction models. *BMC Med.* **17**, 192 (2019).
40. Mattsson, N. et al. Prevalence of the apolipoprotein E ϵ 4 allele in amyloid β positive subjects across the spectrum of Alzheimer's disease. *Alzheimers Dement.* **14**, 913–924 (2018).
41. Jansen, W. J. et al. Prevalence estimates of amyloid abnormality across the Alzheimer disease clinical spectrum. *JAMA Neurol.* **79**, 228–243 (2022).
42. Chatterjee, P. et al. Plasma A β 42/40 ratio, p-tau181, GFAP, and NfL across the Alzheimer's disease continuum: a cross-sectional and longitudinal study in the AIBL cohort. *Alzheimers Dement. J. Alzheimers Assoc.* **19**, 1117–1134 (2023).
43. Lantero-Rodriguez, J. et al. Clinical performance and head-to-head comparison of CSF p-tau235 with p-tau181, p-tau217 and p-tau231 in two memory clinic cohorts. *Alzheimers Res. Ther.* **15**, 48 (2023).
44. Therriault, J. et al. Frequency of biologically defined Alzheimer disease in relation to age, sex, APOE ϵ 4, and cognitive impairment. *Neurology* **96**, e975–e985 (2021).
45. Palmqvist, S. et al. Prediction of future Alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. *Nat. Med.* **27**, 1034–1042 (2021).
46. Petrazzuoli, F. et al. Brief cognitive tests used in primary care cannot accurately differentiate mild cognitive impairment from subjective cognitive decline. *J. Alzheimers Dis.* **75**, 1191–1201 (2020).
47. Bell, C. C. DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. *JAMA* **272**, 828–829 (1994).
48. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn, Text Revision (DSM-5-TR) (American Psychiatric Association, 2022).
49. Borland, E., Stomrud, E., van Westen, D., Hansson, O. & Palmqvist, S. The age-related effect on cognitive performance in cognitively healthy elderly is mainly caused by underlying AD pathology or cerebrovascular lesions: implications for cutoffs regarding cognitive impairment. *Alzheimers Res. Ther.* **12**, 30 (2020).
50. Borland, E. et al. The Montreal Cognitive Assessment: normative data from a large Swedish population-based cohort. *J. Alzheimers Dis.* **59**, 893–901 (2017).

51. Shirk, S. D. et al. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimers Res. Ther.* **3**, 32 (2011).
 52. Levey, A. S. et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney Int.* **80**, 17–28 (2011).
 53. Lundqvist, R. et al. Implementation and validation of an adaptive template registration method for ¹⁸F-flutemetamol imaging data. *J. Nucl. Med.* **54**, 1472–1478 (2013).
 54. Palmqvist, S. et al. Performance of fully automated plasma assays as screening tests for Alzheimer disease-related β -amyloid status. *JAMA Neurol.* **76**, 1060–1069 (2019).
 55. Cullen, N. C. et al. Plasma biomarkers of Alzheimer's disease improve prediction of cognitive decline in cognitively unimpaired elderly populations. *Nat. Commun.* **12**, 3555 (2021).
 56. Blennow, K. et al. Predicting clinical decline and conversion to Alzheimer's disease or dementia using novel Elecsys A β (1–42), pTau and tTau CSF immunoassays. *Sci. Rep.* **9**, 19024 (2019).
 57. Vanderstichele, H. M. J. et al. Optimized standard operating procedures for the analysis of cerebrospinal fluid A β 42 and the ratios of A β isoforms using low protein binding tubes. *J. Alzheimers Dis.* **53**, 1121–1132 (2016).
 58. Hansson, O. et al. The Alzheimer's Association international guidelines for handling of cerebrospinal fluid for routine clinical measurements of amyloid β and tau. *Alzheimers Dement. J. Alzheimers Assoc.* **17**, 1575–1582 (2021).
 59. Gobom, J. et al. Validation of the LUMIPULSE automated immunoassay for the measurement of core AD biomarkers in cerebrospinal fluid. *Clin. Chem. Lab. Med.* **60**, 207–219 (2022).
 60. Hixson, J. E. & Vernier, D. T. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J. Lipid Res.* **31**, 545–548 (1990).
 61. Leitão, M. J. et al. Clinical validation of the Lumipulse G cerebrospinal fluid assays for routine diagnosis of Alzheimer's disease. *Alzheimers Res. Ther.* **11**, 91 (2019).
 62. Therriault, J. et al. Determining amyloid- β positivity using ¹⁸F-AZD4694 PET imaging. *J. Nucl. Med.* **62**, 247–252 (2021).
 63. Ossenkoppele, R. et al. Associations between tau, A β , and cortical thickness with cognition in Alzheimer disease. *Neurology.* **92**, e601–e612 (2019).
 64. Steyerberg, E. W., Eijkemans, M. J. C., Harrell, F. E. Jr & Habbema, J. D. F. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat. Med.* **19**, 1059–1079 (2000).
 65. Harrell, F. E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (Springer International Publishing, 2015).
 66. Steyerberg E. W. *Clinical Prediction Models* (Springer, 2009).
 67. Akaike, H. Likelihood of a model and information criteria. *J. Econom.* **16**, 3–14 (1981).
 68. Harrell, F. E. Jr. rms: regression modeling strategies. R package version 4.1.1 <https://cran.r-project.org/web/packages/rms/index.html> (2023).
- (grant no.2022-1259) and the Swedish federal government under the ALF agreement (grant no. 2022-Projekt0080). W.S.B. is supported by CAPES (grant nos. 88887.372371/2019-00 and 88887.596742/2020-00) and Stiftelsen för Gamla Tjänarinnor. K.B. is supported by the Swedish Research Council (grant no. 2017-00915), the Swedish Alzheimer Foundation (grant nos. AF-930351, AF-939721 and AF-968270), Hjärnfonden, Sweden (grant nos. FO2017-0243 and ALZ2022-0006), the Swedish state under the agreement between the Swedish government and the county councils, the ALF agreement (grant nos. ALFGBG-715986 and ALFGBG-965240) and the Alzheimer's Association 2021 Zenith Award (no. ZEN-21-848495). Doses of [¹⁸F]flutemetamol injection were sponsored by GE Healthcare. The funding sources had no role in the design and conduct of the study, in the collection, analysis and interpretation of the data or in the preparation, review or approval of the manuscript. The TRIAD study is supported by the Weston Brain Institute, Canadian Institutes of Health Research (CIHR) (grant nos. MOP-11-51-31 and RFN 152985, 159815 and 162303), Canadian Consortium of Neurodegeneration and Aging (CCNA; grant no. MOP-11-51-31-team 1), the Alzheimer's Association (grant nos. NIRG-12-92090 and NIRP-12-259245), Brain Canada Foundation (CFI Project no. 34874; 33397), Fonds de Recherche du Québec—Santé (Chercheur Boursier, grant no. 2020-VICO-279314), CCNA (theme 1, team 2) and the Colin J. Adair Charitable Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

W.S.B., K.B. and O.H. designed the study. S.J., J.T., N.R., C.T., J.T., J.S., S.S., E.R.Z., S.P., E.S., P.R.N. and O.H. acquired the clinical data or processed neuroimaging results. S.J., N.J.A., A.L.B., G.T.-B., H.C.K., P.R.N., K.B. and O.H. coordinated and/or performed blood-biomarker quantification. W.S.B., N.C.C. and J.T. performed data analyses. W.S.B., N.C.C., K.B. and O.H. wrote the initial draft of the manuscript. All authors contributed to the interpretation of the results and towards subsequent manuscript drafts.

Funding

Open access funding provided by Lund University.

Competing interests

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 AC Immune, Amylyx, Alzpath, BioArctic, Biogen, Cerveau, Eisai, Eli Lilly, Fujirebio, Genentech, Merck, Novartis, Novo Nordisk, Roche, Sanofi and Siemens. K.B. has served as a consultant, at advisory boards, or at data monitoring committees, for Abcam, Axon, BioArctic, Biogen, JOMDD/Shimadzu, Julius Clinical, Lilly, MagQu, Novartis, Ono Pharma, Pharmatrophix, 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. G.T.-B. and H.C.K. are employees of Janssen Research and Development. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s43587-023-00471-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43587-023-00471-5>.

Correspondence and requests for materials should be addressed to Wagner S. Brum or Oskar Hansson.

Acknowledgements

The work was supported the Swedish Research Council (grant no. 2022-00775), ERA PerMed (grant no. ERAPERMED2021-184), the Knut and Alice Wallenberg Foundation (grant no. 2017-0383), the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation (grant no.AF-980907), the Swedish Brain Foundation (grant no. FO2021-0293), the Parkinson Foundation of Sweden (grant no. 1412/22), the Cure Alzheimer's fund, the Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, the Skåne University Hospital Foundation (grant no. 2020-0000028), Regionalt Forskningsstöd

Peer review information *Nature Aging* thanks Geir Selbaek and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023


Wagner S. Brum ^{1,2} , **Nicholas C. Cullen**³, **Shorena Janelidze**³, **Nicholas J. Ashton**^{1,4,5,6}, **Eduardo R. Zimmer** ^{2,7,8,9}, **Joseph Therriault** ^{10,11}, **Andrea L. Benedet**¹, **Nesrine Rahmouni**^{10,11}, **Cécile Tissot** ^{10,11}, **Jenna Stevenson**^{10,11}, **Stijn Servaes**^{10,11}, **Gallen Triana-Baltzer**¹², **Hartmuth C. Kolb** ¹², **Sebastian Palmqvist** ^{3,13}, **Erik Stomrud**^{3,13}, **Pedro Rosa-Neto** ^{10,11}, **Kaj Blennow** ^{1,14,15} & **Oskar Hansson** ^{3,13,15} 

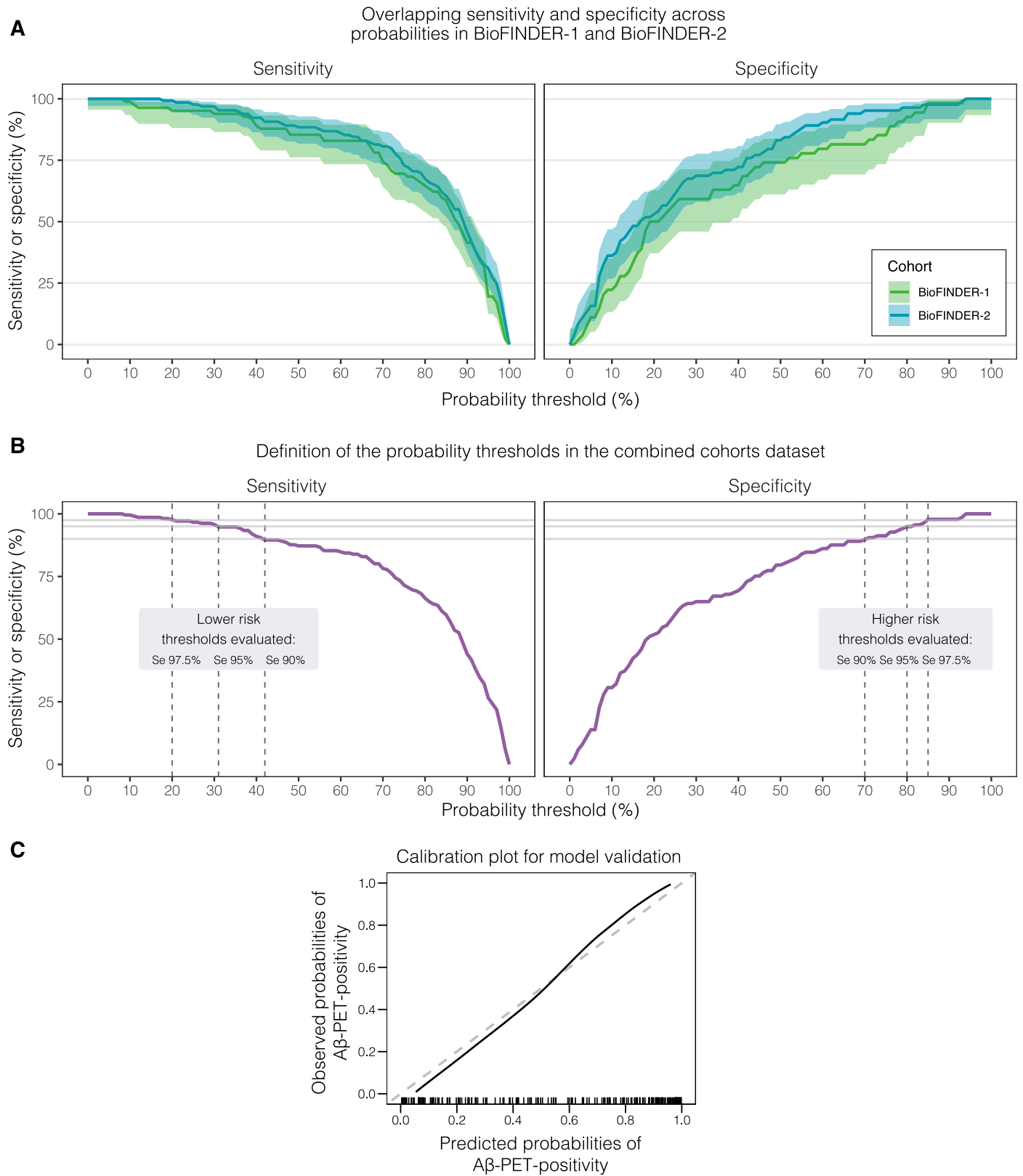
¹Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden. ²Graduate Program in Biological Sciences: Biochemistry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

³Clinical Memory Research Unit, Department of Clinical Sciences, Malmö, Lund University, Lund, Sweden. ⁴Institute of Psychiatry, Psychology and Neuroscience, Maurice Wohl Institute Clinical Neuroscience, King's College London, London, UK. ⁵NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia, South London and Maudsley NHS Foundation, London, UK. ⁶Centre for Age-Related Medicine, Stavanger University Hospital, Stavanger, Norway. ⁷Department of Pharmacology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

⁸Graduate Program in Biological Sciences: Pharmacology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. ⁹McGill Centre for Studies in Aging, McGill University, Montreal, Québec, Canada. ¹⁰Translational Neuroimaging Laboratory, McGill Research Centre for Studies in Aging, Montreal, Québec, Canada. ¹¹Department of Neurology and Neurosurgery, Faculty of Medicine, McGill University, Montreal, Québec, Canada.

¹²Neuroscience Biomarkers, Janssen Research & Development, La Jolla, CA, USA. ¹³Memory Clinic, Skåne University Hospital, Malmö, Sweden.

¹⁴Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden. ¹⁵These authors jointly supervised this work: Kaj Blennow, Oskar Hansson.  e-mail: wagner.scheeren.brum@gu.se; oskar.hansson@med.lu.se

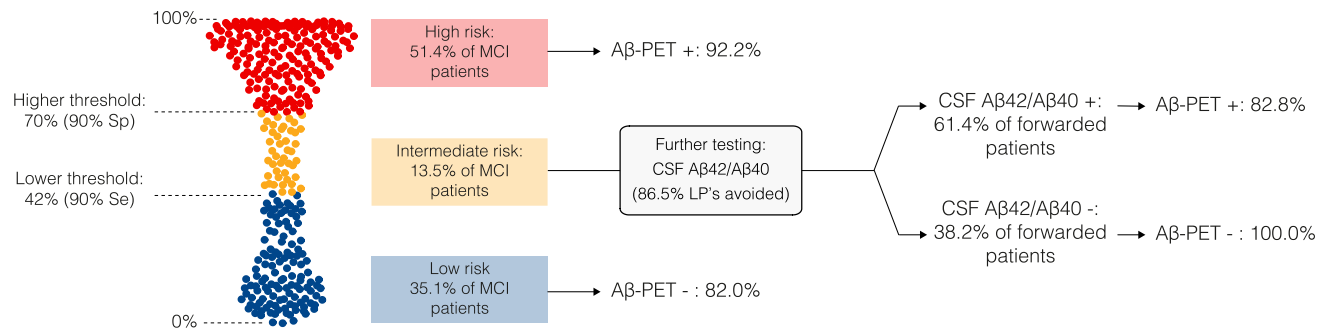


Extended Data Fig. 1 | See next page for caption.

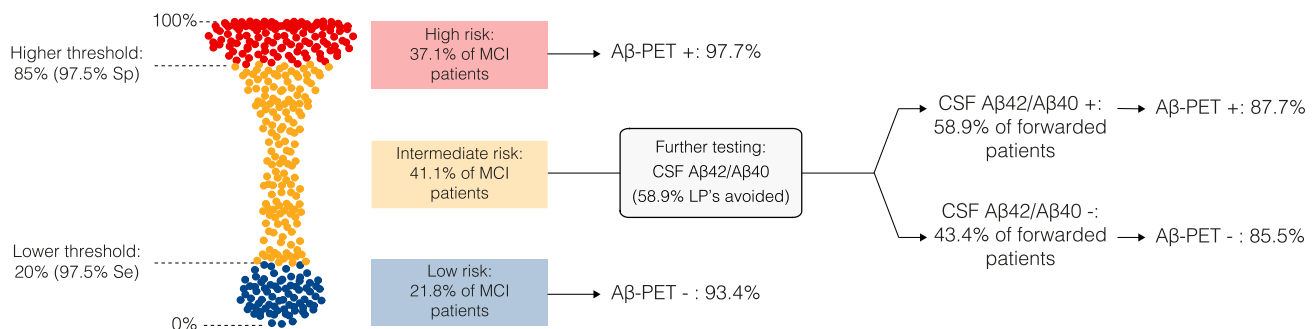
Extended Data Fig. 1 | Sensitivity and specificity across possible probability thresholds in both cohorts and derivation of probability thresholds. (a) Sensitivity and specificity across probability thresholds in BioFINDER-1 and BioFINDER-2, separately presented. The x-axis corresponds to the full range of possible thresholds for the probabilities of A β -PET positivity based on a plasma p-tau217-based model for A β -PET positivity. Solid lines correspond to the observed sensitivities and specificities point estimates for the range of possible probability thresholds, and ribbons to 95% confidence intervals, with BioFINDER-1 in light green and BioFINDER-2 in dark green. **(b)** Since sensitivity and specificity overlapped across the range of possible thresholds in both cohorts, we derived risk stratification thresholds based on predictions from both datasets combined. The lower-risk probability thresholds (left) evaluated

were 42% (resulting in a sensitivity of 90%), 31% (resulting in a sensitivity of 95%) and 20% (resulting in a sensitivity of 97.5%), while the higher-risk probability thresholds (right) evaluated were 70% (resulting in a specificity of 90%), 80% (resulting in a specificity of 95%) and 85% (resulting in a specificity of 97.5%). **(c)** Calibration plot showing external validation in BioFINDER-2 of the model derived in BioFINDER-1. The solid black line shows smoothed associations between the predicted probabilities and observed frequencies of A β -PET positivity. The closer this line is to the dotted grey identity line, the better performing and more generalizable a prediction model is. A β = Amyloid- β . PET = Positron emission tomography. P-tau217 = tau phosphorylated at threonine 217. Se = Sensitivity. Sp = Specificity.

A - CSF testing following plasma biomarker risk stratification (90% Sensitivity/Specificity thresholds)

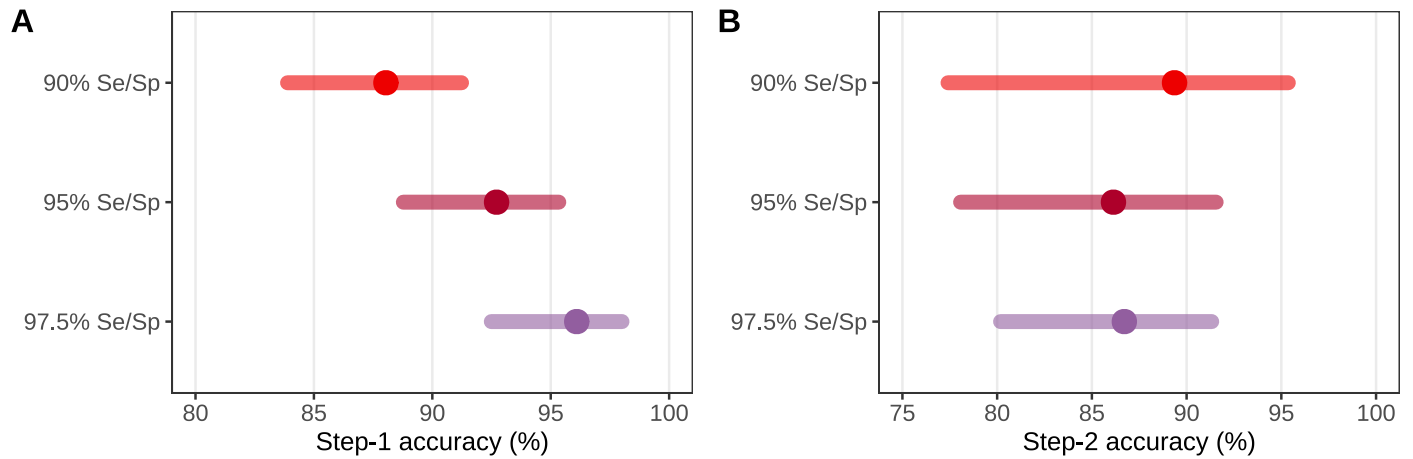


B - CSF testing following plasma biomarker risk stratification (97.5% Sensitivity/Specificity thresholds)



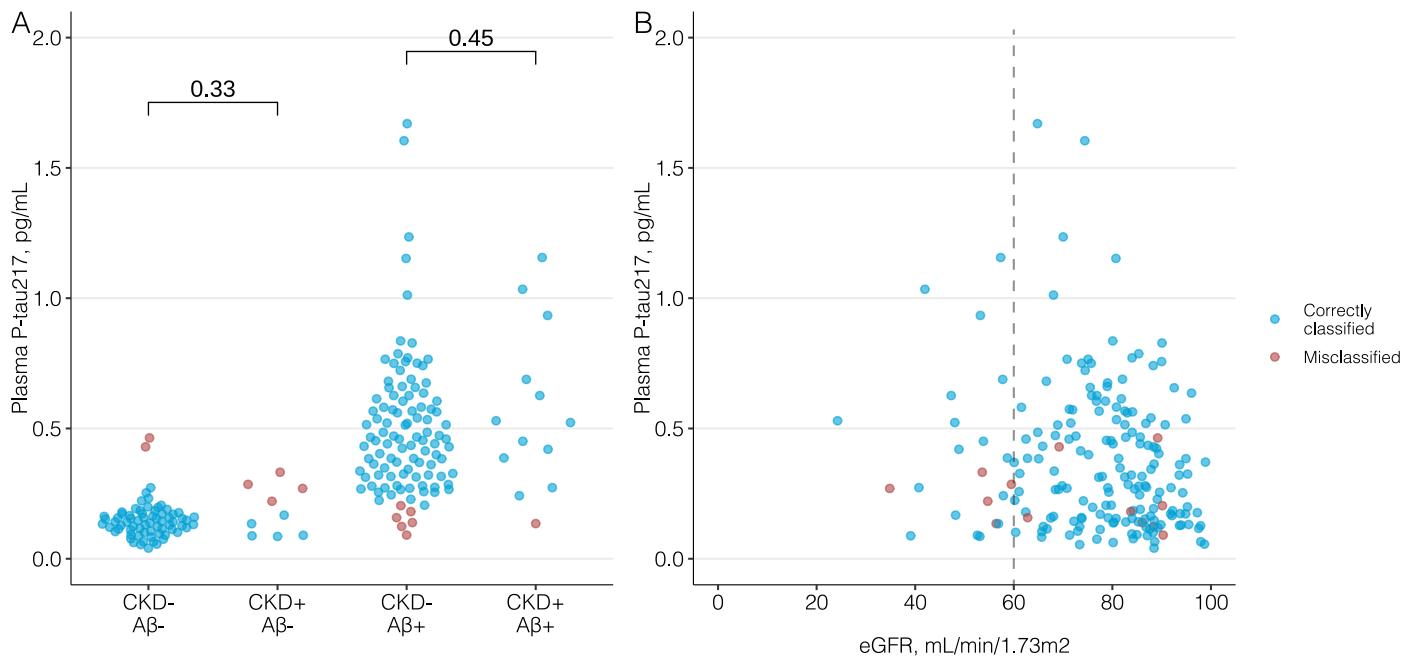
Extended Data Fig. 2 | Flowchart illustrating plasma-based risk stratification and further testing of intermediate-risk individuals with CSF Aβ42/Aβ40 for the 90 and 97.5% strategies. Flowchart recapitulating results from the first step of the workflow (blood biomarker-based risk stratification) and demonstrating the accuracy for the second step of the clinical workflow, when intermediate-risk individuals are referred to CSF Aβ42/Aβ40 test for predicting Aβ-PET status.

(a) Shows results for the 90% Se/Sp risk stratification strategy, and **(b)** for the 97.5% Se/Sp strategy. The 95% Se/Sp strategy is represented in Fig. 1b of the main text. Aβ = Amyloid-β. PET = Positron emission tomography. CSF = cerebrospinal fluid. P-tau217 = tau phosphorylated at threonine 217. Se = Sensitivity. Sp = Specificity. LP = lumbar puncture.



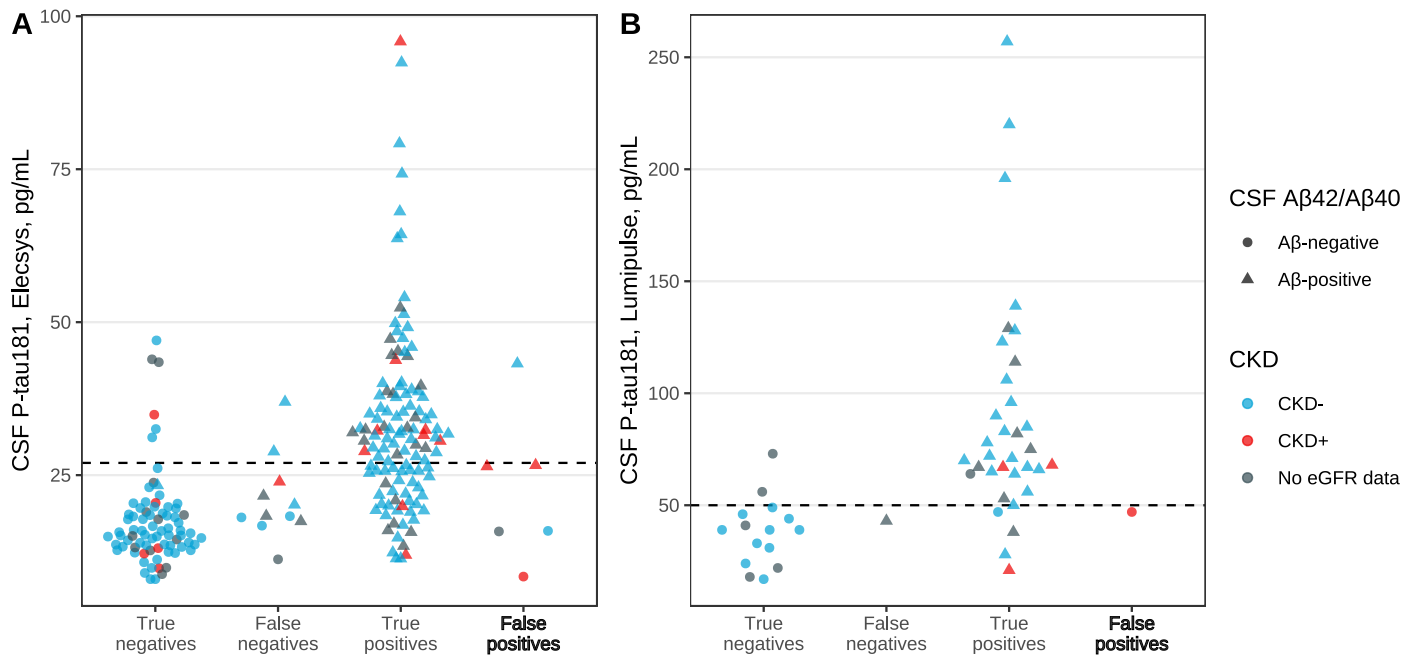
Extended Data Fig. 3 | Separate-step performance. For each of the graphics, the x-axis corresponds to the three evaluated strategies for blood-based biomarker risk stratification (Se/Sp 90%; Se/Sp 95%; Se/Sp 97.5%), with dots representing point estimates and bars corresponding to 95% confidence intervals, computed for the BioFINDER-1 and BioFINDER-2 combined population ($n = 348$). **(a)** Indicates the overall accuracy for the low- and high-risk groups for the workflow's first step, that is blood-based biomarker risk stratification. This metric was calculated based on the number of A β -PET negative individuals classified to the low-risk group and of A β -PET positive individuals classified to

the high-risk group (90% Se/Sp: $n = 265$; 95% Se/Sp: $n = 229$; 97.5% Se/Sp: $n = 197$), divided by the total individuals in the high and low-risk groups. **(b)** Displays the accuracy for the second step of the workflow. Individuals in the intermediate-risk group were assumed to be forwarded to a lumbar puncture to test CSF A β 42/40 test, and the accuracy corresponding to the overall concordance of a CSF-negative result with a negative A β -PET scan and of a CSF-positive results with a positive A β -PET scan (90% Se/Sp: $n = 42$; 95% Se/Sp: $n = 87$; 97.5% Se/Sp: $n = 143$). A β = Amyloid- β . PET = Positron emission tomography. CSF = cerebrospinal fluid. P-tau217 = tau phosphorylated at threonine 217. Se = Sensitivity. Sp = Specificity.



Extended Data Fig. 4 | Reduced renal function does not seem to influence plasma p-tau217 levels between Aβ-PET negative and positive individuals classified as low- or high-risk. (a) The dots represent plasma p-tau217 concentrations (y-axis), with the x-axis representing the Aβ-PET status in combination with chronic kidney disease (CKD) as determined by an eGFR below 60 mL/min/1.73m². Only participants classified as low- or high-risk at step-1 of the workflow based on the plasma p-tau217 model-derived probabilities and the 95% Se/Sp strategy are included. The colors indicate whether patients were correctly classified (blue) or misclassified (red). P-values come from t-tests (two-sided, alpha 0.05) were used to assess whether plasma p-tau217 levels were altered by the presence of CKD among Aβ-negatives and Aβ-positive participants. Plasma p-tau217 levels did not were not significantly altered by

CKD among Aβ-negative participants or among Aβ-positive participants. **(b)** The y-axis represents plasma p-tau217 levels and the x-axis represents continuous values of eGFR, with the with the colors indicating whether patients were correctly classified (blue) or misclassified (red) at the step-1 of the workflow based on the plasma p-tau217 model-derived probabilities and the 95% Se/Sp strategy. The plot shows that misclassifications occur throughout all the span of renal function. Further, the dashed line indicates most of the misclassified individuals with CKD were, in fact, very close to the eGFR cutoff for CKD of 60 mL/min/1.73m². Aβ = Amyloid-β. PET = Positron emission tomography. eGFR = estimated glomerular filtration rate. CKD = chronic kidney disease. P-tau217 = tau phosphorylated at threonine 217. Se = Sensitivity. Sp = Specificity.



Extended Data Fig. 5 | Patients with CKD misclassified as false-positives often presented high CSF p-tau levels or CSF-positivity for Aβ42/Aβ40. This figure represents CSF p-tau181 levels and CSF Aβ42/Aβ40 status measured for BioFINDER-1 and BioFINDER-2 patients according to their classification status at the 95% Se/Sp risk stratification strategy with the main analysis plasma p-tau217-based model. The y-axis and dots displays CSF p-tau181, with colors representing CKD status (CKD-, blue; CKD+, red) and the shapes correspond to CSF Aβ42/Aβ40 status. In the x-axis, patients are stratified into true-negatives (low-risk label at step-1 who were also Aβ-PET-negative), false-negatives (low-risk label at step-1 who were also Aβ-PET-positive), true-positives (high-risk label at step-1 who were also Aβ-PET-positive), false-positives (high-risk label at step-1 who were Aβ-PET-negative), with intermediate-risk individuals excluded from the plot (assumed to be referred for a CSF test with no applicable correct/incorrect classification label). **(a)** Displays CSF biomarker results measured with Elecsys for BioFINDER-1 and most of BioFINDER-2 patients, with the horizontal line corresponding to a previously validated cut-off for p-tau181 of 28 pg/mL (ref. 56).

(b) Displays CSF biomarker results measured with Lumipulse a subset of BioFINDER-2 patients, with the horizontal line corresponding to a previously validated cut-off for p-tau181 of 50.2 pg/mL (ref. 59). For both assays, CSF Aβ42/Aβ40 was handled as described in the methods. Given the false-positive group ($n = 7$; x-axis, in bold) had demonstrated a higher rate of CKD after classification with a plasma p-tau217 risk stratification model, this figure indicates that $n = 3$ out of the $n = 4$ false-positives with CKD had elevated CSF p-tau181 levels (very close to indicated clinical cutoffs), with $n = 2$ of these patients also being positive for CSF Aβ42/Aβ40. This suggests a peripheral increase in plasma p-tau217 in the absence of Aβ-PET-positivity could be related to an underlying disease process (since CSF changes might occur earlier than PET) rather than peripheral impaired clearance. Aβ = Amyloid-β. CSF = cerebrospinal fluid. PET = Positron emission tomography. CKD = chronic kidney disease. P-tau181 = tau phosphorylated at threonine 181. P-tau217 = tau phosphorylated at threonine 217. Se = Sensitivity. Sp = Specificity.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

This study does not include data available in external or online repositories. Anonymized data will be shared by request from a qualified academic investigator for the sole purpose of replicating procedures and results presented in the article. For BioFINDER, requests will be considered as long as data transfer is in agreement with EU legislation on the general data protection regulation and decisions by the Swedish Ethical Review Authority and Region Skåne, which should be regulated in

a material transfer agreement, and contact can be made through the study's website (<https://biofinder.se/>). Arrangements for data sharing for replication of the findings in the TRIAD data set are subject to standard data-sharing agreements, and further information can be found in the study's website (<https://triad.tn-mcgill.com/>).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	We used the term "sex" in the manuscript, and sex was determined based on participant self-reporting. The cohorts included recruit both self-reported men and women.
Population characteristics	Detailed information on demographics from the two cohorts, BioFINDER-1 and BioFINDER-2, is provided in Table 1. Participants had mild cognitive impairment (MCI) and had available data for biomarkers relevant for this study (A β -PET, CSF A β 42/A β 40, plasma p-tau217, APOE genotype). Overall, the study population (n=348) had a median (IQR) age of 73 (63.-76.5) years, 39.4% of participants were female, 52.9% were APOE4 carriers, and median (IQR) years of educational attainment were 12 (9-14). Common comorbidities in older adults were also present, such as cardiovascular disease (54.0%), diabetes (15.8%), and dyslipidemia (37.9%). In TRIAD, participants were slightly younger [mean (SD) age 68 (8) years], with slightly more females (n=46; 55%), with key demographic characteristics summarised in Supplementary Table 6.
Recruitment	BioFINDER-1 (NCT01208675) and BioFINDER-2 (NCT03174938) are memory clinic based cohorts recruiting participants forwarded from primary care to secondary memory clinics in southern Sweden due to cognitive complaints. BioFINDER-1 participants were recruited between January 2010 and January 2015. BioFINDER-2 started recruitment in May 2017. TRIAD is a cohort enrolling patients mostly from memory clinic settings, as well as volunteers. In both cohorts, memory clinic patients are invited by their care providers to participate in the study. The potential bias of these recruitment strategies is that the findings will be more restricted to memory patients, with a limited generalizability to primary care. It is discussed in the manuscript that such primary care validation is warranted by further studies.
Ethics oversight	All BioFINDER and TRIAD patients gave their written informed consent to participate in the study, and participation was voluntary. The BioFINDER studies were approved by the Ethical Review Board in Lund, Sweden, which is part of the Swedish Ethical Review Authority (BioFINDER-1: 2010-156; BioFINDER-2: 2016-1053). TRIAD was approved by the Montreal Neurological Institute PET working committee and the Douglas Mental Health University Institute Research Ethics Board (IUSMD16-61, IUSMD16-60).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Our development cohort (BioFINDER-1: n=136) and validation cohorts (BioFINDER-2: n=212; TRIAD: n=84) were cross-sectionally analyzed based on complete biomarker availability for plasma, genetic, cerebrospinal fluid and imaging data. There is no indication that we were insufficiently powered for these analyses considering a potential predictive utility of these algorithms. Obtaining such a number of individuals with such a comprehensive biomarker panel is a strength of the study, and similar clinical prediction modeling AD studies, that included patients with a more restrict biomarker panel, had roughly similar numbers (Verberk et al <i>Ann Neurol</i> , 2018 Nov;84(5):648-658; Janelidze et al <i>Alzheimers Dement</i> , 2022 Feb;18(2):283-293; Hu et al, <i>JAMA Netw Open</i> , 2022 Apr 1;5(4):e228392).
Data exclusions	The main inclusion criteria was that participants presented cognitive impairment, relevant clinical phenotype for current detection of AD pathology for considering potential novel treatments. Also, participants needed to have available data, for the same baseline visit, for plasma p-tau217, cerebrospinal fluid A β 42/A β 40, and A β -PET. Both BioFINDER-1 and BioFINDER-2 have recruited participants with MCI and dementia. However, per protocol in both cohorts, A β -PET is not carried in the later clinical stage of dementia. Thus, our analyses focused on the MCI population with the required biomarkers for testing our hypothesis. Within the eligible participants, no datapoint or individual was excluded from any analyses.
Replication	When developing a prediction model, such as we did in BioFINDER-1, validating it externally (i.e. obtaining risk probabilities using the coefficients from the developed model) it is crucial to determine its potential utility. The model developed in BioFINDER-1 demonstrated to have validated very well in BioFINDER-2, displaying very high discrimination and good calibration, with this validation being tested only once. For the geographical validation, the model also validated well, and an independent analyst (JT) carried data analyses, with validation being performed only once as well.
Randomization	No allocation into experimental groups was performed. Thus, randomization is not relevant to this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Included in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Plasma phosphorylated tau 217 (p-tau217) was quantified in both BioFINDER cohorts with an immunoassay developed by Lilly Research Laboratories, and analyses were performed with the same batch of reagents. Biotinylated-IBA493 was used as a capture antibody and SULFO-TAG-4G102 as the detector. In TRIAD, plasma p-tau217 was quantified with an immunoassay developed by Janssen R&D, using the PT3 antibody as capture and HT43 as detector. Samples and antibodies in both assays were diluted in 1:2. These antibodies are research-use only and are not commercialized.

Validation

The plasma p-tau217 immunoassays here used has been previously widely validated for this application and described in full detail (Lilly immunoassay: Palmqvist et al., JAMA. 2020;324(8):772-781; Janssen immunoassay: Groot et al, Alz Res Ther; 2022 May 14;14(1):67).

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

BioFINDER-1: NCT01208675. BioFINDER-2: NCT03174938.

Study protocol

BioFINDER-1: <https://clinicaltrials.gov/ct2/show/NCT01208675>. BioFINDER-2: <https://clinicaltrials.gov/ct2/show/NCT03174938>.

Data collection

BioFINDER-1 participants were recruited between January 2010 and January 2015. BioFINDER-2 started recruitment in May 2017.

Outcomes

The primary outcome of the study was the overall accuracy of the two-step workflow for A β -status, composed of a first step with a blood-biomarker prediction model, and further testing with CSF A β 42/40 for individuals with uncertain results at the first step. This outcome would inform on how accurate classification was in general, for both steps. Another primary outcome was the number of further confirmatory tests reduced by step-1. Interpreted these two outcomes together enables the evaluation of the trade-off between accuracy and number of confirmatory tests, which is the main focus of the paper. A β -PET-status, the reference standard involved in the outcome, was defined by 18-F-flutemetamol PET, using previously validated cutoffs.