



# Prediction of future Alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures

Sebastian Palmqvist<sup>1,2</sup>✉, Pontus Tideman<sup>1,2</sup>, Nicholas Cullen<sup>1</sup>, Henrik Zetterberg<sup>3,4,5,6</sup>, Kaj Blennow<sup>3,4</sup>, the Alzheimer's Disease Neuroimaging Initiative\*, Jeffery L. Dage<sup>7</sup>, Erik Stomrud<sup>1,2</sup>, Shorena Janelidze<sup>1</sup>, Niklas Mattsson-Carlgren<sup>1,8,9</sup> and Oskar Hansson<sup>1,2</sup>✉

**A combination of plasma phospho-tau (P-tau) and other accessible biomarkers might provide accurate prediction about the risk of developing Alzheimer's disease (AD) dementia. We examined this in participants with subjective cognitive decline and mild cognitive impairment from the BioFINDER ( $n = 340$ ) and Alzheimer's Disease Neuroimaging Initiative (ADNI) ( $n = 543$ ) studies. Plasma P-tau, plasma A $\beta$ 42/A $\beta$ 40, plasma neurofilament light, APOE genotype, brief cognitive tests and an AD-specific magnetic resonance imaging measure were examined using progression to AD as outcome. Within 4 years, plasma P-tau217 predicted AD accurately (area under the curve (AUC) = 0.83) in BioFINDER. Combining plasma P-tau217, memory, executive function and APOE produced higher accuracy (AUC = 0.91,  $P < 0.001$ ). In ADNI, this model had similar AUC (0.90) using plasma P-tau181 instead of P-tau217. The model was implemented online for prediction of the individual probability of progressing to AD. Within 2 and 6 years, similar models had AUCs of 0.90–0.91 in both cohorts. Using cerebrospinal fluid P-tau, A $\beta$ 42/A $\beta$ 40 and neurofilament light instead of plasma biomarkers did not improve the accuracy significantly. The clinical predictions by memory clinic physicians had significantly lower accuracy (4-year AUC = 0.71). In summary, plasma P-tau, in combination with brief cognitive tests and APOE genotyping, might greatly improve the diagnostic prediction of AD and facilitate recruitment for AD trials.**

Correctly determining if a patient with subtle cognitive symptoms, such as memory decline, suffers from prodromal or preclinical AD and will progress to AD dementia within the near future remains a challenge for clinicians. The task is nonetheless of utmost importance for a timely referral to a memory clinic, a correct and early AD diagnosis, initiation of symptomatic treatment, planning for the future and, hopefully soon, for initiating disease-modifying treatments. Although there have been impressive developments in biomarkers for AD and progression to AD dementia, such as cerebrospinal fluid (CSF) analysis of  $\beta$ -amyloid (A $\beta$ 42 (ref. 1)) or the ratio of A $\beta$ 42/A $\beta$ 40 (ref. 2), P-tau<sup>3,4</sup> and neurofilament light (NFL)<sup>5</sup>, as well as A $\beta$ -positron emission tomography (PET)<sup>6,7</sup> and tau-PET<sup>8,9</sup>, the invasive nature, high cost and limited availability restrict their use to a limited number of highly specialized centers. A possible turning point has emerged with the recent development of blood-based biomarkers, making it possible to measure NFL<sup>10,11</sup>, A $\beta$ 42/A $\beta$ 40 (refs. 12,13) and P-tau (phosphorylated at either threonine 181 or threonine 217) in plasma<sup>14–16</sup>.

Plasma P-tau181 and P-tau217, in particular, have shown especially high diagnostic performance for discriminating AD dementia from other neurodegenerative diseases<sup>14–16</sup>. Plasma P-tau has also recently been shown to be suitable for individualized prediction of cognitive decline in individuals with mild cognitive impairment (MCI)<sup>17</sup>. In the clinical workup of patients with cognitive complaints,

however, it is unlikely that plasma P-tau (or any other biomarker) will achieve the highest potential predictive accuracy on its own, owing to the multifactorial nature of AD etiology and its heterogeneous clinical presentation. There is, therefore, now a need to identify which other measures plasma P-tau should be combined with to produce the most accurate prediction of future AD and establish an optimal diagnostic algorithm of non-invasive, cost-effective and easily available methods for early diagnosis of AD.

Furthermore, before establishing plasma P-tau in clinical practice, alone or as part of an algorithm, it is important to examine whether it actually performs better than the clinical prediction made by a treating physician, which has not been previously examined. The aim of the current study was, therefore, to examine the accuracy of plasma P-tau among patients with mild cognitive symptoms for predicting future AD dementia when combined with other accessible and non-invasive biomarkers. The prediction included not only the discrimination between progression to AD dementia and stable cognitive symptoms but also versus progression to other dementias. The accuracies were compared with the diagnostic prediction of memory clinic physicians who had performed extensive clinical assessments and evaluated cognitive testing and structural brain imaging at baseline. Selection of variables and accuracies from the models were examined in two independent, multi-center cohorts. The primary cohort was the Swedish BioFINDER study,

<sup>1</sup>Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Lund, Sweden. <sup>2</sup>Memory Clinic, Skåne University Hospital, Malmö, Sweden. <sup>3</sup>Department of Psychiatry and Neurochemistry, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden. <sup>4</sup>Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden. <sup>5</sup>Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK. <sup>6</sup>UK Dementia Research Institute at University College London, London, UK. <sup>7</sup>Eli Lilly and Company, Indianapolis, IN, USA. <sup>8</sup>Department of Neurology, Skåne University Hospital, Lund, Sweden. <sup>9</sup>Wallenberg Center for Molecular Medicine, Lund University, Lund, Sweden. \*A list of authors appears at the end of the paper. ✉e-mail: [sebastian.palmqvist@med.lu.se](mailto:sebastian.palmqvist@med.lu.se); [oskar.hansson@med.lu.se](mailto:oskar.hansson@med.lu.se)

**Table 1 | Baseline characteristics for the BioFINDER and ADNI cohorts**

	BioFINDER			ADNI		
	Converting to AD dementia (n = 91)	Not converting to AD dementia (n = 249)	P value	Converting to AD dementia (n = 102)	Not converting to AD dementia (n = 441)	P value
Age (years)	72.1 (4.91)	70.2 (5.73)	<b>0.007</b>	73.2 (6.98)	71.2 (7.14)	<b>0.003</b>
Gender (female)	44 (48%)	124 (50%)	0.81	54 (53%)	229 (52%)	0.85
Education (years)	11.8 (3.44)	11.9 (3.56)	0.75	16.2 (2.64)	16.3 (2.63)	0.64
Baseline diagnosis						
SCD	20 (22.0%)	144 (57.8%)	<b>&lt;0.001</b>	0 (0%)	106 (24.0%)	<b>&lt;0.001</b>
MCI	71 (78.0%)	105 (42.2%)	<b>&lt;0.001</b>	102 (100%)	335 (76.0%)	<b>&lt;0.001</b>
Time to any dementia from baseline SCD (years)	3.54 (1.71)	3.39 (1.26)		NA	3.76 (2.49)	
Time to any dementia from baseline MCI (years)	2.68 (1.52)	2.85 (1.49)		2.54 (1.90)	3.16 (2.42)	
Follow-up time, stable participants (years)	NA	4.55 (1.62)		NA	4.30 (2.38)	
Follow-up diagnosis						
AD	91 (100%)	0		102 (100%)	0	
DLB/PDD/PSP	0	10 (4.0%)		0	1 (0.2%)	
FTD	0	7 (2.8%)		0	1 (0.2%)	
VaD	0	23 (9.2%)		0	2 (0.5%)	
Other	0	8 (3.2%)		0	21 (4.8%) <sup>a</sup>	
Non-dementia	0	201 (80.7%)		0	416 (94.3%)	
MMSE (points)	27.1 (1.87)	28.2 (1.65)	<b>&lt;0.001</b>	27.4 (1.81)	28.5 (1.56)	<b>&lt;0.001</b>
Memory function (z-score)	-0.70 (0.82)	0.25 (0.94)	<b>&lt;0.001</b>	-0.82 (1.01)	0.20 (0.89)	<b>&lt;0.001</b>
Verbal function (z-score)	-0.096 (0.89)	0.035 (0.75)	0.11	-0.39 (1.17)	0.090 (0.93)	<b>&lt;0.001</b>
Executive function (z-score)	-0.46 (0.74)	0.17 (0.87)	<b>&lt;0.001</b>	-0.493 (0.927)	0.114 (0.777)	<b>&lt;0.001</b>
Visuospatial function (z-score)	-0.17 (0.96)	0.061 (1.01)	0.02	-0.33 (1.26)	0.075 (0.91)	<b>0.002</b>
Plasma P-tau217 (pg ml <sup>-1</sup> )	0.40 (0.25)	0.17 (0.14)	<b>&lt;0.001</b>	NA	NA	
Plasma P-tau181 (pg ml <sup>-1</sup> ) <sup>b</sup>	4.20 (2.22)	2.26 (4.54)	<b>&lt;0.001</b>	24.4 (10.8)	15.8 (11.4)	<b>&lt;0.001</b>
Plasma Aβ42/Aβ40 (pg ml <sup>-1</sup> )	0.062 (0.006)	0.066 (0.008)	<b>&lt;0.001</b>	NA	NA	
Plasma NFL (pg ml <sup>-1</sup> ) <sup>c</sup>	25.2 (10.9)	23.6 (20.8)	<b>0.002</b>	44.9 (18.1)	35.1 (19.0)	<b>&lt;0.001</b>
Cortical thickness of the AD signature region (mm) <sup>c</sup>	2.30 (0.24)	2.48 (0.27)	<b>&lt;0.001</b>	2.65 (0.18)	2.76 (0.15)	<b>&lt;0.001</b>

Data are shown as mean (s.d.) or n (%). Group comparisons were performed using the Mann-Whitney U test. <sup>a</sup>There were 17 participants with a clinical follow-up diagnosis of AD dementia who were Aβ-negative. These were coded as having other dementias. <sup>b</sup>Note that plasma P-tau181 was available in only a small subset in BioFINDER, and that pre-analytical and assay differences make the BioFINDER concentrations incomparable with the concentrations in ADNI. <sup>c</sup>Calibration-related (NFL) or camera-related (MRI) differences make it difficult to directly compare the results between the cohorts for these biomarkers. Note that, before analyzed in logistic regression models, biomarker concentrations were transformed so that higher values corresponded to more abnormal results. DLB, dementia with Lewy bodies; FTD, frontotemporal degeneration; NA, not applicable; PDD, pervasive developmental disorder; PSP, progressive supranuclear palsy; VaD, vascular dementia.

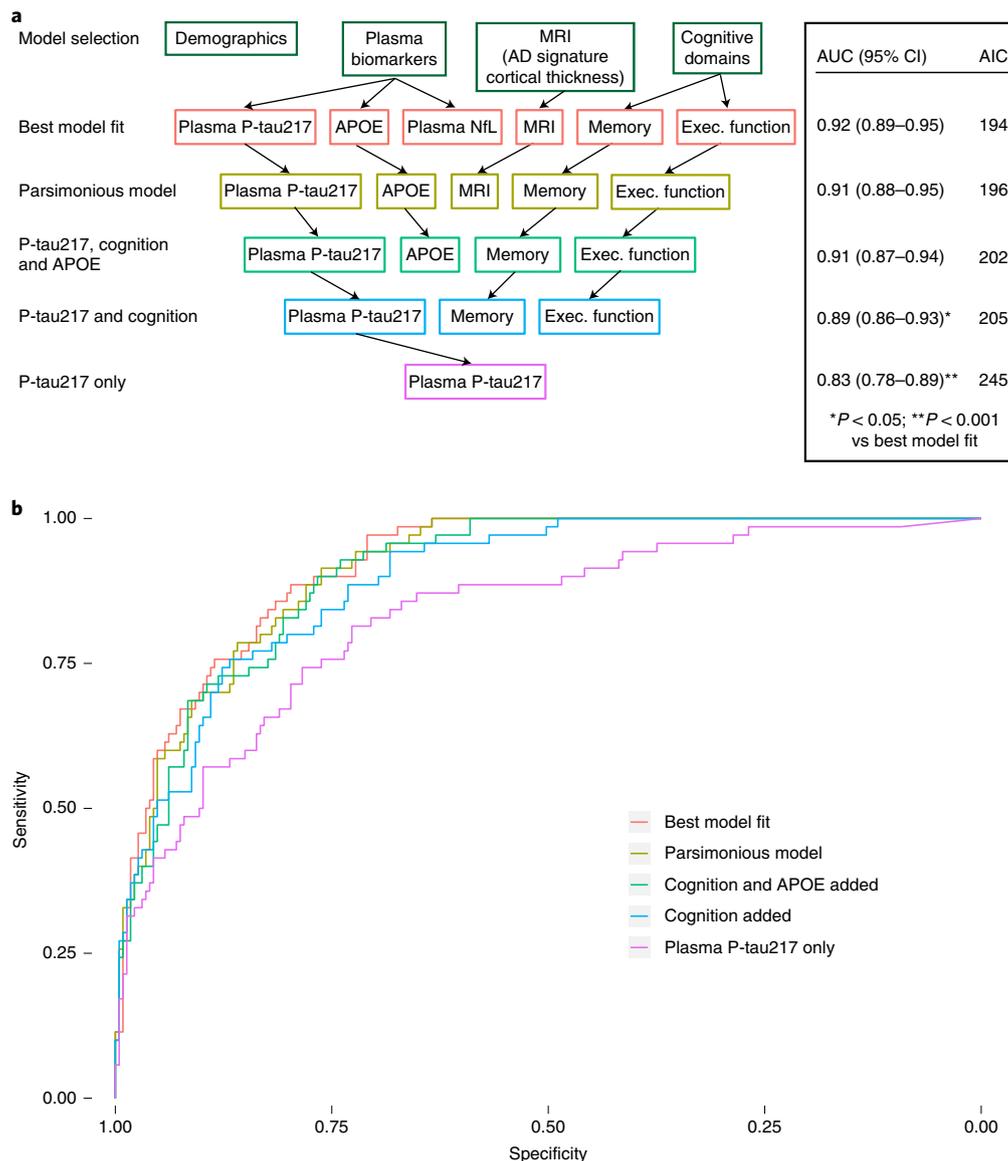
and the validation cohort was the ADNI. The primary outcome was progression to AD dementia within 4 years; secondary outcomes were progression to AD dementia within 2 and 6 years, respectively. Finally, a cross-validated model was established and implemented as an online tool for predicting the individual risk of progressing to AD dementia (<http://predictAD.app>).

## Results

**Participants in BioFINDER.** The cohort included 340 consecutively enrolled patients with cognitive complaints. One hundred and sixty-four patients were subsequently characterized as having subjective cognitive decline (SCD), and 176 patients were subsequently characterized as having MCI. Ninety-one patients progressed to AD dementia at follow-up; 48 patients progressed to other dementias; and 201 patients did not progress to any dementia. The mean (s.d.) age was 70.7 (5.6) years, and 49% were women. The mean (s.d.) time

to dementia was 2.9 (1.5) years, and the mean (s.d.) follow-up time in participants who did not progress to dementia was 4.5 (1.6) years. Participant characteristics are described in Table 1; the enrollment process is described in Extended Data Fig. 1.

**Prediction of AD dementia in BioFINDER.** Figure 1 summarizes the model selection process and main results. First, a data-driven model selection was performed to select the model with the lowest Akaike information criterion (AIC)—that is, the best tradeoff between model fit and model complexity for predicting AD dementia (see Supplementary Methods for a detailed description of AIC). Variables screened for included key demographics, number of *APOE* ε4 alleles, brief tests from four cognitive domains, a magnetic resonance imaging (MRI) measure ('AD signature' cortical thickness from temporal regions prone to atrophy in AD<sup>18</sup>) and plasma biomarkers (NFL, P-tau217 and Aβ42/Aβ40). Then, a parsimonious

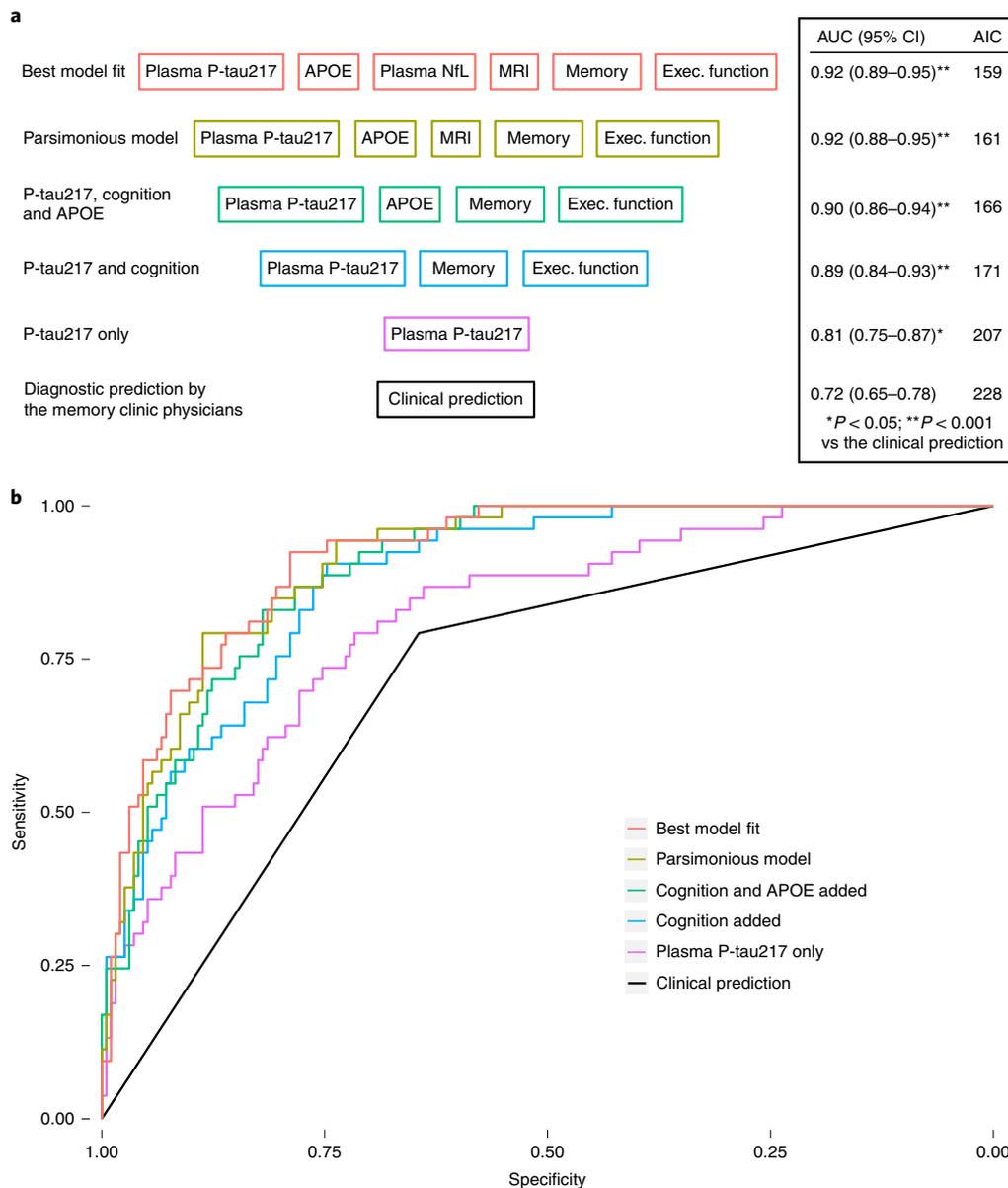


**Fig. 1 | Model selection process and performance for predicting AD dementia within 4 years. a**, The logistic regression model selection process. Best model fit shows the data-driven model selection with the lowest AIC (that is, the model with best fit). The parsimonious model shows the model that had a similar model performance ( $\Delta AIC < 2$ ) with as few significant predictors as possible. In subsequent models, modalities were removed in a stepwise procedure. Model specifications, including comparisons between all models, are shown in Supplementary Table 1. Comparisons between AUCs were performed using DeLong statistics. **b**, ROC curve analyses of the different models for discriminating those who progressed to AD dementia versus those who progressed to other dementias or remained cognitively stable.  $n = 297$ , of whom 70 progressed to AD dementia within the time period. *APOE*, apolipoprotein E genotype (number of  $\epsilon 4$  alleles); Exec. function, executive function; MRI, magnetic resonance imaging of the cortical thickness of a temporal AD-specific region; ROC, receiver operating characteristic.

model was created by removing as many variables as possible while maintaining a similar model performance defined as being within two AIC points of the lowest AIC model identified in the first step ( $\Delta AIC < 2$ )<sup>19,20</sup>. Thereafter, variables were removed further in a stepwise procedure to examine the performance of more basic models. Using the primary outcome—prediction of AD dementia within 4 years—the best model included the predictors plasma P-tau217, number of *APOE*  $\epsilon 4$  alleles, executive function, memory function, cortical thickness and plasma NfL (Fig. 1 and Supplementary Table 1). This model resulted in an AUC of 0.92 (95% confidence interval (CI) 0.89–0.95). Removing plasma NfL resulted in similar model performance ( $\Delta AIC < 2$ ) and accuracy (AUC = 0.91, 95% CI 0.88–0.95). Also removing cortical thickness (that is, a model with plasma

P-tau217, cognition and *APOE*) retained a similar accuracy (AUC = 0.91, 95% CI 0.87–0.94,  $P = 0.09$ –0.82 versus the more complex models, including the model with best fit) but with slightly poorer model fit ( $\Delta AIC 6.0$ –7.7). The AUC of the latter model was significantly higher than when using only plasma P-tau217 (AUC = 0.83, 95% CI 0.78–0.89,  $P < 0.001$ ; Fig. 1 and Supplementary Table 1).

The model with best fit for predicting AD dementia within 2 years resulted in similarly included variables and accuracies (Extended Data Fig. 3 and Supplementary Table 2). Model selection for prediction within 6 years selected the same variables as for the 4-year outcome but with the addition of plasma A $\beta 42$ /A $\beta 40$  (AUC = 0.94, 95% CI 0.91–0.97; Extended Data Fig. 4 and Supplementary Table 3).



**Fig. 2 | Comparison with the clinical prediction for predicting AD dementia within 4 years. a**, Logistic regression models from Fig. 1 compared with the clinical diagnostic prediction of memory clinic physicians at the baseline visit (before fluid biomarker data were available). Comparisons between AUCs were performed using DeLong statistics. **b**, ROC curve analyses of the different models for discriminating those who progressed to AD dementia versus those who progressed to other dementias or remained cognitively stable. Note that the comparison with the clinical prediction was performed on a subsample where the clinical prediction was available, hence the slightly different AUCs (and 95% CIs) compared with those shown in Fig. 1.  $n = 247$ , of whom 53 progressed to AD dementia within the time period. APOE, apolipoprotein E genotype (number of  $\epsilon 4$  alleles); Exec. function, executive function; MRI, magnetic resonance imaging of the cortical thickness of a temporal AD-specific region; ROC, receiver operating characteristic.

Separate AUCs of univariable models for predicting progression to AD dementia at each time point from 2 to 6 years are shown and discussed in Extended Data Fig. 2.

**The clinical prediction of AD dementia in BioFINDER.** The subsample where the memory clinic physicians at baseline determined the most probable underlying cause of the cognitive impairment (that is, ‘clinical prediction’) comprised 285 patients, of whom 72 converted to AD dementia during follow-up. Using the primary outcome (progression to AD dementia within 4 years), the AUC for the clinical prediction was 0.72 (95% CI 0.65–0.78), and, for plasma P-tau217 alone, the AUC was 0.81 (95% CI 0.75–0.87) ( $P = 0.03$

versus the clinical prediction). Adding memory, executive function and APOE to plasma P-tau217 provided a further significantly higher accuracy than the clinical prediction (AUC = 0.90, 95% CI 0.86–0.94,  $P < 0.001$ ) (see Fig. 2 and Supplementary Table 4 for details). Significantly better accuracies for the models versus the clinical prediction were also seen at 2 and 6 years (Supplementary Tables 5 and 6).

**Comparison with CSF biomarkers.** To examine the effect of using CSF biomarkers instead of plasma biomarkers, we tested different models with CSF P-tau,  $A\beta_{42}/A\beta_{40}$  and NfL for the main outcome: prediction of AD dementia within 4 years (Supplementary Table 7).

In the model with best fit (Fig. 1, top model), the use of CSF P-tau and CSF NFL, instead of plasma P-tau and plasma NFL, produced a non-significantly different AUC (0.93, 95% CI 0.90–0.96,  $P=0.44$ ). In a new data-driven model selection using CSF instead of plasma biomarkers, CSF P-tau, CSF A $\beta$ 42/A $\beta$ 40, memory, executive function and cortical thickness were selected for the model with best fit. This model also produced a non-significantly different AUC when compared to the model with best fit using plasma biomarkers (0.94, 95% CI 0.92–0.97,  $P=0.085$ ). The more basic model with P-tau, *APOE*, memory and executive function had very similar accuracies (AUC = 0.91 using either CSF or plasma P-tau;  $P=0.55$ ). Finally, in univariate analyses, both CSF and plasma P-tau had an AUC of 0.83 ( $P=0.95$ ).

**Validation in the ADNI cohort.** The cohort included 106 participants with SCD and 437 participants with MCI, of whom 102 progressed to AD dementia at follow-up and 28 to other dementias (Table 1). The validation from BioFINDER was carried out in two steps. First, the same type of model selection was performed in ADNI to examine if similar variables were selected (with the exception that plasma P-tau181 was available instead of P-tau217 and that plasma A $\beta$ 42/A $\beta$ 40 measures were available only in a small subsample and, therefore, not included in the analysis;<sup>21,22</sup> see Supplementary Methods for details). Second, the key models identified from the BioFINDER cohort were tested in ADNI. Third, a cross-validated model was constructed and implemented online (see next section).

For the primary outcome (predicting AD dementia within 4 years), the same biomarkers as in BioFINDER were selected for the model with best fit in ADNI, with the exception that plasma NFL was not chosen in ADNI (Fig. 3 and Supplementary Table 8). Note that, even though plasma NFL was selected in BioFINDER, it was not a significant predictor (Supplementary Table 1). When testing the variables selected in the parsimonious model from BioFINDER in ADNI (plasma P-tau, MRI, *APOE*, memory and executive function), the accuracy (AUC = 0.91, 95% CI 0.87–0.94 in BioFINDER) was not different from the model with best fit established in ADNI (AUC = 0.91, 95% CI 0.87–0.94,  $P=0.41$ ,  $\Delta\text{AIC}<2$ ) (Fig. 3 and Supplementary Table 8). The more basic model with just P-tau, *APOE*, memory and executive function performed very similarly in both cohorts (AUC = 0.90, 95% CI 0.86–0.94 in ADNI versus AUC = 0.91, 95% CI 0.87–0.94 in BioFINDER).

Similar accuracies were seen in ADNI compared to BioFINDER for predictions of AD dementia within 2 and 6 years, respectively (Supplementary Table 2 versus Supplementary Table 9 and Supplementary Table 3 versus Supplementary Table 10).

Finally, a comparison between models using CSF P-tau181 instead of plasma P-tau181 was performed. The plasma-based models did not perform significantly differently from the CSF-based models (Supplementary Table 11). For example, plasma P-tau181, *APOE*, memory and executive function had an AUC of 0.90 (95% CI 0.86–0.94), and CSF P-tau181, *APOE*, memory and executive function had an AUC of 0.91 (95% CI 0.88–0.95). When comparing plasma P-tau181 alone (AUC = 0.78, 95% CI 0.72–0.84) to CSF P-tau181 alone (AUC = 0.86, 95% CI 0.81–0.90), plasma had a lower accuracy ( $P=0.01$ ).

#### Cross-validation and implementation of a prediction algorithm.

As seen in the comparison between the BioFINDER and ADNI cohorts, a model consisting of plasma P-tau, memory, executive function and number of *APOE*  $\epsilon$ 4 alleles provided a good balance among simplicity, accuracy and generalizability for prediction of AD dementia within 4 years (AUCs 0.90–0.91; Figs. 1 and 4). We, therefore, created a new model in BioFINDER, where the estimates of the model could be directly tested in ADNI (and other cohorts). Because of the different plasma P-tau isoforms used in the

cohorts, P-tau217 and P-tau181 were converted to binary variables (abnormal/normal). Unbiased plasma P-tau cutoffs were established in independent populations in BioFINDER and ADNI at the mean level + 2 s.d. in A $\beta$ -negative controls (Methods). Cognitive domain scores for both the BioFINDER and ADNI cohorts were established based on the distribution in an independent control sample in BioFINDER without tau or A $\beta$  pathology, adjusted for age and education (Supplementary Methods)<sup>23,24</sup>. The new model had an AUC of 0.89 (95% CI 0.85–0.93) in BioFINDER ( $P=0.18$  versus the model using continuous plasma P-tau217 data in Fig. 1). When the estimates of the model were validated in ADNI, the AUC was 0.86 (95% CI 0.81–0.90) (Fig. 4a). A secondary cross-validated model was constructed using plasma P-tau z-scores (instead of binary P-tau status) based on the distribution in A $\beta$ -negative control participants (Methods). This model had an AUC of 0.90 (95% CI 0.87–0.94) in BioFINDER, and, when the estimates of the model were validated in ADNI, the AUC was very similar (0.89, 95% CI 0.85–0.93) (Extended Data Fig. 5). The cross-validated models were implemented online at <http://predictAD.app> where the individual probability of progressing to AD dementia within 4 years can be calculated for new cases (Fig. 4b).

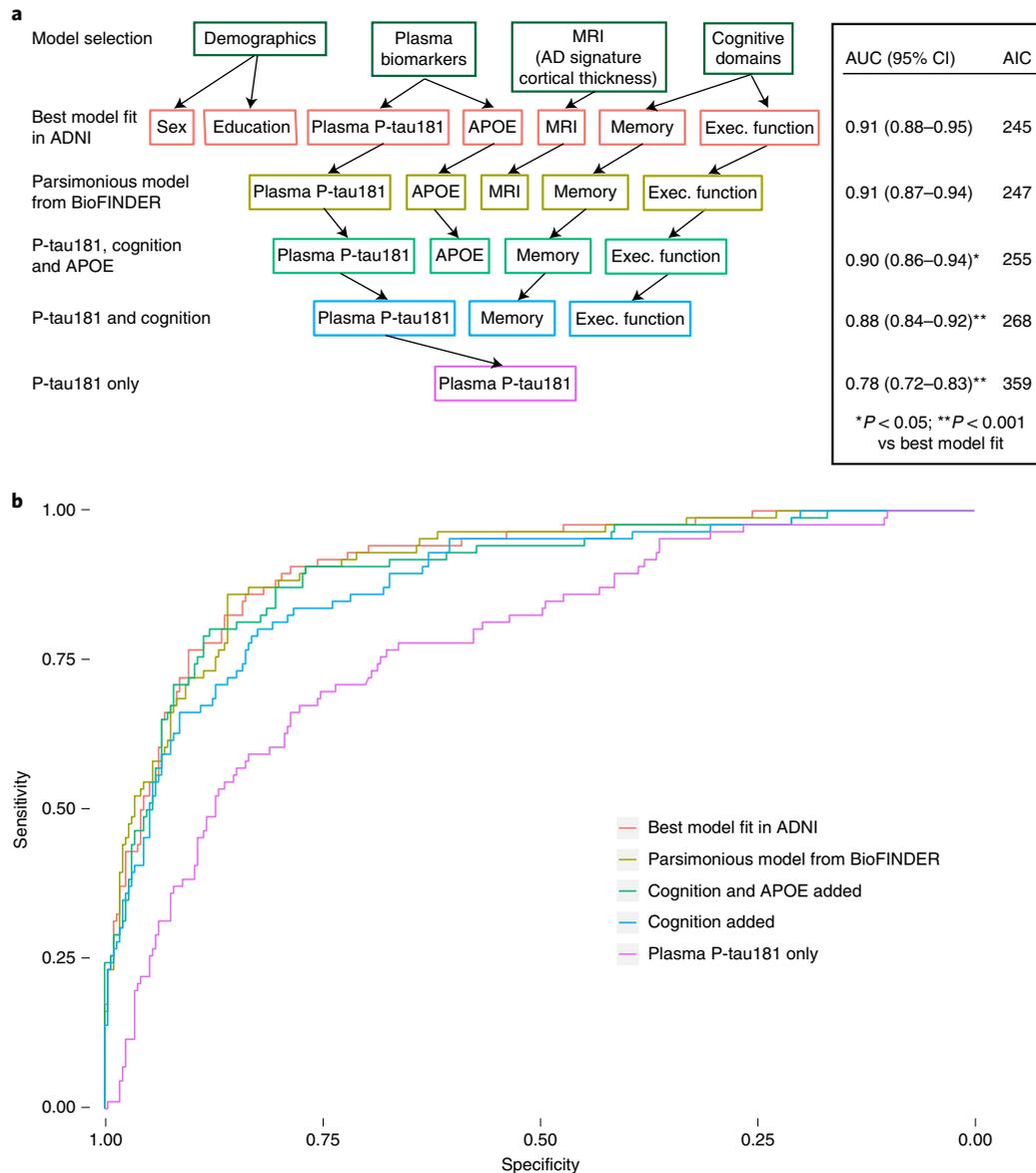
**Additional analyses.** In primary care centers and in centers not specifically specialized in dementia disorders, patients without dementia are unlikely to be correctly classified as having SCD or MCI, because it requires a substantial cognitive battery<sup>25</sup>. However, this is often done in specialized memory clinics, and, for this sake, separate analyses in patients with SCD and MCI were performed. When evaluating the predictive accuracy in patients with SCD, we found that plasma P-tau217, *APOE*, memory and executive function could predict development of AD dementia within 4 years with an AUC of 0.95 in BioFINDER (Supplementary Table 12). When selecting only individuals with MCI, we found that plasma P-tau217, *APOE*, memory and executive function could predict AD dementia within 4 years with an AUC of 0.86, which increased to 0.88 when also including MRI in BioFINDER (Supplementary Table 14). In ADNI, the same variables could predict AD dementia in MCI with an AUC of 0.90 (Supplementary Table 17). Details of these and additional analyses in SCD and MCI populations are shown in Supplementary Tables 12–19.

In addition, negative and positive predictive values for the main models in the whole populations of BioFINDER and ADNI are reported in Supplementary Tables 20 and 21.

Finally, a plasma P-tau181 measure (although with a different assay and analytical platform than for the one used in ADNI) was available in a subsample in BioFINDER ( $n=192$  for the 4-year prediction). When comparing plasma P-tau217 to this P-tau181 measure for predicting progression to AD dementia within 4 years, no significant differences in predictive accuracies were found (Supplementary Table 22).

#### Discussion

In this study, we examined how plasma P-tau could be best combined with other easily accessible and cost-effective measures to predict progression to AD dementia, primarily within 4 years, in a heterogenous and consecutively included memory clinic cohort. Although plasma P-tau alone could predict AD dementia accurately within 4 years (AUC = 0.83), the most marked increase in accuracy was seen when it was combined with brief cognitive tests of memory and executive function and *APOE* genotype (AUC = 0.91; Fig. 1 and Supplementary Table 1). Minor further improvements were seen when also including cortical thickness and plasma NFL (Fig. 1). No significant differences were observed in accuracies when using CSF biomarkers instead of plasma biomarkers (Supplementary Table 7 and 11). Plasma P-tau217 alone and in combination with other variables had significantly higher accuracy



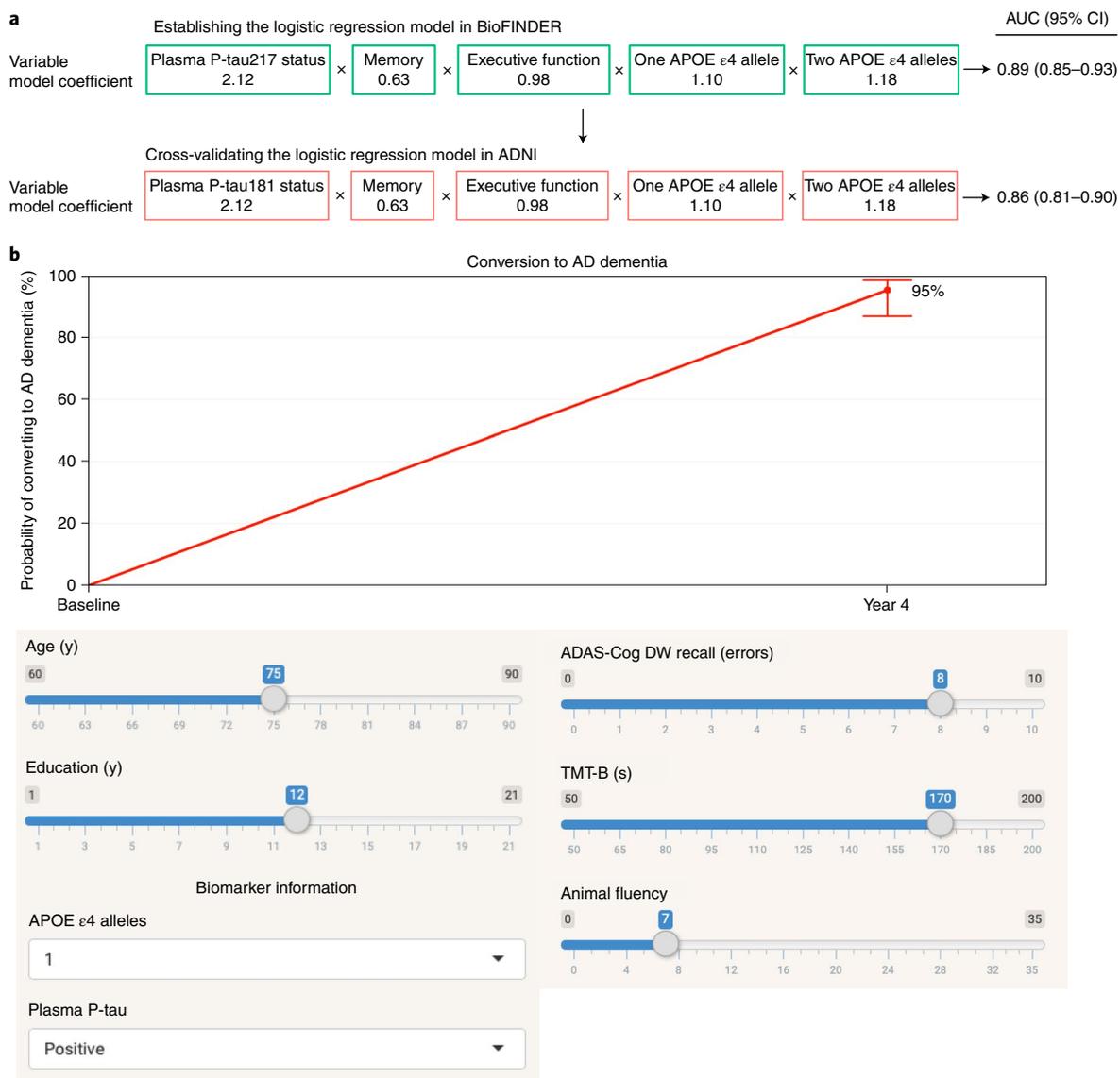
**Fig. 3 | Model selection and performance in ADNI for predicting AD dementia within 4 years with comparisons using models selected in BioFINDER.**

**a**, Logistic regression model selection process. Best model fit shows the data-driven model selection with the lowest AIC in the ADNI cohort. This was compared to the models selected in BioFINDER and applied in ADNI. Model specifications are shown in Supplementary Table 8. Comparisons between AUCs were performed using DeLong statistics. **b**, ROC curve analyses of the different models for discriminating those who progressed to AD dementia versus those who progressed to other dementias or remained cognitively stable.  $n = 376$ , of whom 86 progressed to AD dementia within the time period. APOE, apolipoprotein E genotype (number of  $\epsilon 4$  alleles); Exec. function, executive function; MRI, magnetic resonance imaging of the cortical thickness of a temporal AD-specific region; ROC, receiver operating characteristic.

than the clinical diagnostic predictions of memory clinic physicians after a comprehensive baseline assessment including medical history, cognitive testing and computed tomography (CT) or MRI of the brain (Fig. 2 and Supplementary Table 4). The generalizability of these predictors was demonstrated by a similar variable selection and performance in the independent ADNI cohort (Fig. 3) as well as the performance in separate SCD and MCI populations, respectively (Supplementary Tables 12–19). Especially the combination of plasma P-tau, memory, executive function and APOE genotype had a robust performance and high accuracy in both cohorts (AUC = 0.90–0.91 in BioFINDER and ADNI; Figs. 1 and 4). This selection of variables was used to create cross-validated models that can be used to predict the individual risk for progression

to AD dementia for new cases (Fig. 4 and Extended Data Fig. 5; <http://predictAD.app>).

Although there has been great progress recently in validating plasma P-tau as a biomarker for AD<sup>14–16,26–30</sup> and for individualized prediction of cognitive decline in individuals classified as having MCI<sup>17</sup>, this paper is, to our knowledge, the first to present how plasma P-tau can be combined with other easily available and cost-effective measures for predicting development of AD dementia in patients seeking medical care based on diverse mild cognitive symptoms. To ensure robustness of the models that included plasma P-tau and other measures, the other measures that were examined in this study were based on the literature of cognitive tests sensitive to the cognitive decline in AD<sup>31–33</sup>, plasma and MRI biomarkers that



**Fig. 4 | Cross-validation and implementation of an algorithm.** **a**, Cross-validation of the logistic regression model using plasma P-tau status instead of continuous plasma P-tau levels. Model coefficients were established in BioFINDER (AUC = 0.89) and tested in ADNI (AUC = 0.86). z-scores were inverted so that higher scores equal poorer results. A secondary cross-validated model using continuous plasma P-tau z-scores was also established in BioFINDER (AUC = 0.90) and tested in ADNI (AUC = 0.89) (Extended Data Fig. 5). BioFINDER:  $n = 297$ , of whom 70 progressed to AD dementia within the time period. ADNI:  $n = 376$ , of whom 86 progressed to AD dementia within the time period. **b**, Implementation of the logistic regression models at <http://predictAD.app>, where one can enter the raw cognitive test scores, number of APOE  $\epsilon$ 4 alleles and plasma P-tau status (from either P-tau217 or P-tau181). Alternatively, a plasma P-tau z-score can be entered for a higher predictive accuracy. Age and education are not part of the logistic regression model but are used to calculate cognitive z-scores. The example shows that the predicted individual risk of progressing to AD dementia within 4 years is 95% (error bars represent the 95% prediction interval) in a 75-year old individual with cognitive complaints who has 12 years of education, one APOE  $\epsilon$ 4 allele and abnormal plasma P-tau status and scores eight errors on a ten-word delayed recall test (that is, remembers two out of ten words), names seven animals in 1 min and completes the Trail-Making Test B (TMT-B) in 170 s. ADAS-Cog DW recall, 10-word delayed recall test from the ADAS-Cog.

have been shown to measure the underlying disease processes in AD at different stages<sup>11,12,16,18</sup> and known demographic and genetic risk factors of AD and cognitive decline<sup>34,35</sup>. Another novelty of the study is the comparison with the clinical prediction (Fig. 2), which shows the true value of implementing plasma P-tau alone or in combination with other measures to improve the diagnostic prediction in clinical practice.

Using BioFINDER as the primary cohort of interest had some valuable strengths for determining optimal combinations of tests for use in clinical practice. The population consisted of consecutively recruited patients who had been referred to participating memory

clinics, making the cohort heterogenous and representative of a future target population (Table 1). Nonetheless, similar results were obtained in ADNI that consists of a selected population focused on AD. The cognitive span in both cohorts ranged from subjective to objective cognitive symptoms (that is, both SCD and MCI), which best mimics the clinical scenario where physicians would use the combination of tests. Although the commonly used division into individuals with MCI and cognitively unimpaired individuals (SCD and controls) can make sense from a research standpoint—for example, for studying disease mechanisms in AD<sup>36</sup>—this cognitive classification system is difficult to replicate among cohorts and

translate into clinical practice. Depending on MCI definitions and use of cognitive tests and cutoffs, those classified as MCI will greatly vary between populations<sup>37–39</sup>. And even if a unified definition and set of cognitive tests were determined, the comprehensive cognitive test battery needed would not fit the testing routines—for example, in primary care<sup>25</sup>.

This study had limitations. First, plasma A $\beta$ 42/A $\beta$ 40 was not available in an adequately large sample size in ADNI and could not, therefore, be included in that cohort. However, in BioFINDER, where plasma A $\beta$ 42/A $\beta$ 40 was available, it was selected only for the prediction within 6 years, and removing it from the model reduced the AUC by less than 0.02. Second, the cognitive tests that were available in both cohorts were limited. This resulted in sub-optimal tests for the verbal and visuospatial domains, which could have contributed to their lower accuracy (Extended Data Fig. 2). On the other hand, those representing the memory and executive domains have been extensively validated as sensitive measures of early cognitive decline in AD<sup>31–33</sup>. Third, although we did not find a significantly different accuracy between the best-performing plasma-based models (AUC = 0.92) and CSF-based models (AUC = 0.94), we cannot rule out that using a much larger sample size would have identified a significant difference. Fourth, the updated diagnostic criteria for AD are based on a framework of identifying A $\beta$  (A), tau (T) and, depending on stage of the disease, neurodegeneration (N)<sup>36</sup>. From this ATN scheme, we included only a biomarker for T (P-tau) in our online algorithm consisting of plasma P-tau, *APOE* and cognition. However, in this model, we think that the number of *APOE*  $\epsilon$ 4 alleles acts as a proxy for A $\beta$  given their incrementally increased association with A $\beta$  burden<sup>34</sup>. As shown when using CSF instead of plasma biomarkers, CSF A $\beta$ 42/40 (which is a better biomarker for A $\beta$  pathology than plasma A $\beta$ 42/40), and not *APOE* genotype, was selected in the model with the lowest AIC (Supplementary Table 7, model 3). In addition, P-tau has also been shown to reflect A $\beta$  pathology (that is, it is not a pure tau biomarker)<sup>4,40</sup>, further supporting that the predictors, including in the algorithm, partly reflect both A and T in the ATN scheme. Regarding ‘N’, although it is an important marker for determining the pathophysiological stage of AD, the inclusion of cognitive tests might have overruled its importance in predicting cognitive decline, because cognition itself probably is more closely linked to progression to dementia in individuals who already experience cognitive symptoms. Fifth, a model using estimates based on absolute plasma P-tau concentrations could not be cross-validated, because different P-tau isoforms were available in BioFINDER (plasma P-tau217) and ADNI (plasma P-tau181). Instead, a cross-validated model including binary plasma P-tau (abnormal/normal), memory, executive function and number of *APOE*  $\epsilon$ 4 alleles was validated (Fig. 4). This model had high accuracy in the training cohort but, more importantly, also when applied in the validation cohort. Furthermore, we developed a second cross-validated model using continuous z-scores of plasma P-tau based on a reference population, which resulted in higher accuracy in the cross-validation analysis (Extended Data Fig. 5; both models are implemented at <http://predictAD.app>). In the near future, it is likely that several different P-tau assays will be available on the market. The robust cross-validated results found in the present study, despite using different plasma P-tau assays in BioFINDER versus ADNI, opens up the possibility to use the same algorithm for different P-tau assays with similar prognostic information, including high-performing assays for either plasma P-tau181 or P-tau217, provided that either binary or standardized continuous P-tau data are used.

As for the potential diagnostic improvements in clinical practice, our comparison with the clinical-based prediction shows a clear advantage of using plasma P-tau in combination with the other measures (increases in AUC from 0.72 to 0.89–0.92; Fig. 2). The clinical prediction consisted of the baseline assessment of memory

clinic physicians, showing the potential improved value at a specialist center. In addition, the presented models showed similarly high accuracy when comparing CSF biomarkers and plasma biomarkers (Supplementary Tables 7 and 11). This suggests that the plasma P-tau models might provide a similar substitute for CSF analyses in settings where these are not accessible or too expensive. Venous puncture (for plasma analyses) is also easier for patients to undergo than lumbar puncture (for CSF analysis). In primary care, the implementation of these models is even more important because of the restricted availability of accurate diagnostic tools and the fact that only 20–50% of patients with dementia are routinely recognized and documented<sup>41,42</sup>. Presuming that primary care physicians make less accurate predictions of future AD dementia than memory clinic physicians, the advantage of using brief diagnostic algorithms based on plasma P-tau in primary care would be even greater. However, the present cross-validation does not warrant an accurate prediction of AD dementia in primary care, and further validation is required in large, unselected and ethnically diverse primary care populations with a lower pre-test probability of underlying AD. Additionally, future steps before clinical implementation of plasma P-tau-based prognostic algorithms include (1) development of clinical-grade plasma P-tau assays approved by the appropriate regulatory authorities; (2) establishment of a standardized protocol for collecting and handling of plasma; and (3) agreement on appropriate use criteria to avoid misinterpretation and misuse of these prognostic algorithms in the clinic.

Besides clinical use, a prognostic algorithm could be used for recruitment of individuals with early AD to clinical trials. The presented models might, therefore, provide substantial cost benefits compared to using CSF analysis or PET to screen for eligible participants, which could speed up recruitment and, hence, facilitate the drug development process of future disease-modifying treatments for AD<sup>32</sup>.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-021-01348-z>.

Received: 22 December 2020; Accepted: 12 April 2021;

Published online: 24 May 2021

### References

- Palmqvist, S. et al. Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid  $\beta$ -amyloid 42: a cross-validation study against amyloid positron emission tomography. *JAMA Neurol.* **71**, 1282–1289 (2014).
- Janelidze, S. et al. CSF A $\beta$ 42/A $\beta$ 40 and A $\beta$ 42/A $\beta$ 38 ratios: better diagnostic markers of Alzheimer disease. *Ann. Clin. Transl. Neurol.* **3**, 154–165 (2016).
- Barthelemy, N. R. et al. Cerebrospinal fluid phospho-tau T217 outperforms T181 as a biomarker for the differential diagnosis of Alzheimer’s disease and PET amyloid-positive patient identification. *Alzheimers Res. Ther.* **12**, 26 (2020).
- Mattsson, N. et al. A $\beta$  deposition is associated with increases in soluble and phosphorylated tau that precede a positive tau PET in Alzheimer’s disease. *Sci. Adv.* **6**, eaaz2387 (2020).
- Mattsson-Carlgen, N. et al. The implications of different approaches to define AT(N) in Alzheimer disease. *Neurology* **94**, e2233–e2244 (2020).
- Mattsson, N., Palmqvist, S., Stomrud, E., Vogel, J. & Hansson, O. Staging  $\beta$ -amyloid pathology with amyloid positron emission tomography. *JAMA Neurol.* <https://doi.org/10.1001/jamaneurol.2019.2214> (2019).
- Rabinovici, G. D. et al. Association of amyloid positron emission tomography with subsequent change in clinical management among Medicare beneficiaries with mild cognitive impairment or dementia. *JAMA* **321**, 1286–1294 (2019).
- Leuzy, A. et al. Diagnostic performance of RO948 F 18 tau positron emission tomography in the differentiation of Alzheimer disease from other neurodegenerative disorders. *JAMA Neurol.* **77**, 955–965 (2020).

9. Ossenkuppe, R. et al. Discriminative accuracy of [<sup>18</sup>F]flortaucipir positron emission tomography for Alzheimer disease vs other neurodegenerative disorders. *JAMA* **320**, 1151–1162 (2018).
10. Gisslen, M. et al. Plasma concentration of the neurofilament light protein (NFL) is a biomarker of CNS injury in HIV infection: a cross-sectional study. *EBioMedicine* **3**, 135–140 (2016).
11. Preische, O. et al. Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease. *Nat. Med.* **25**, 277–283 (2019).
12. Palmqvist, S. et al. Performance of fully automated plasma assays as screening tests for Alzheimer disease-related  $\beta$ -amyloid status. *JAMA Neurol.* **76**, 1060–1069 (2019).
13. Schindler, S. E. et al. High-precision plasma  $\beta$ -amyloid 42/40 predicts current and future brain amyloidosis. *Neurology* **93**, e1647–e1659 (2019).
14. Janelidze, S. et al. Plasma P-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia. *Nat. Med.* **26**, 379–386 (2020).
15. 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).
16. Palmqvist, S. et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *JAMA* **324**, 772–781 (2020).
17. Cullen, N. C. et al. Individualized prognosis of cognitive decline and dementia in mild cognitive impairment based on plasma biomarker combinations. *Nat. Aging* **1**, 114–123 (2021).
18. Jack, C. R. Jr. et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement.* **13**, 205–216 (2017).
19. Burnham, K. P. & Anderson, D. R. Multimodel inference: understanding AIC and BIC in model selection. *Soc. Methods Res.* <https://doi.org/10.1177/0049124104268644> (2004).
20. Olofson, E. & Dahan, A. Using Akaike's information theoretic criterion in mixed-effects modeling of pharmacokinetic data: a simulation study. *PLoS ONE* **8**, 71 (2013).
21. Toledo, J. B. et al. Factors affecting A $\beta$  plasma levels and their utility as biomarkers in ADNI. *Acta Neuropathol.* **122**, 401–413 (2011).
22. Ovod, V. et al. Amyloid  $\beta$  concentrations and stable isotope labeling kinetics of human plasma specific to central nervous system amyloidosis. *Alzheimers Dement.* **13**, 841–849 (2017).
23. Borland, E. et al. The Montreal Cognitive Assessment: normative data from a large Swedish population-based cohort. *J. Alzheimers Dis.* **59**, 893–901 (2017).
24. 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).
25. 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).
26. Barthelemy, N. R., Horie, K., Sato, C. & Bateman, R. J. Blood plasma phosphorylated-tau isoforms track CNS change in Alzheimer's disease. *J. Exp. Med.* **217**, e20200861 (2020).
27. Karikari, T. K. et al. Diagnostic performance and prediction of clinical progression of plasma phospho-tau181 in the Alzheimer's Disease Neuroimaging Initiative. *Mol. Psychiatry* **26**, 429–442 (2020).
28. Mattsson-Carlsson, N. et al. Longitudinal plasma p-tau217 is increased in early stages of Alzheimer's disease. *Brain* **143**, 3234–3241 (2020).
29. Mielke, M. M. et al. Plasma phospho-tau181 increases with Alzheimer's disease clinical severity and is associated with tau- and amyloid-positron emission tomography. *Alzheimers Dement.* **14**, 989–997 (2018).
30. Palmqvist, S. et al. Cerebrospinal fluid and plasma biomarker trajectories with increasing amyloid deposition in Alzheimer's disease. *EMBO Mol. Med.* **11**, e11170 (2019).
31. Insel, P. S. et al. Determining clinically meaningful decline in preclinical Alzheimer disease. *Neurology* **93**, e322–e333 (2019).
32. Palmqvist, S. et al. Accurate risk estimation of  $\beta$ -amyloid positivity to identify prodromal Alzheimer's disease: cross-validation study of practical algorithms. *Alzheimers Dement.* **15**, 194–204 (2019).
33. Papp, K. V., Rentz, D. M., Orlovsky, I., Sperling, R. A. & Mormino, E. C. Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: the PACC5. *Alzheimers Dement.* **3**, 668–677 (2017).
34. Jansen, W. J. et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. *JAMA* **313**, 1924–1938 (2015).
35. Scheyer, O. et al. Female sex and Alzheimer's risk: the menopause connection. *J. Prev. Alzheimers Dis.* **5**, 225–230 (2018).
36. Jack, C. R. Jr. et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* **14**, 535–562 (2018).
37. Petersen, R. C. Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* **256**, 183–194 (2004).
38. Roberts, R. O. et al. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology* **30**, 58–69 (2008).
39. Gothlin, M., Eckerstrom, M., Rolstad, S., Wallin, A. & Nordlund, A. Prognostic accuracy of mild cognitive impairment subtypes at different cut-off levels. *Dement. Geriatr. Cogn. Disord.* **43**, 330–341 (2017).
40. Sato, C. et al. Tau kinetics in neurons and the human central nervous system. *Neuron* **97**, 1284–1298 (2018).
41. Lopponen, M., Raiha, I., Isoaho, R., Vahlberg, T. & Kivela, S. L. Diagnosing cognitive impairment and dementia in primary health care—a more active approach is needed. *Age Ageing* **32**, 606–612 (2003).
42. Valcour, V. G., Masaki, K. H., Curb, J. D. & Blanchette, P. L. The detection of dementia in the primary care setting. *Arch. Intern. Med.* **160**, 2964–2968 (2000).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## the Alzheimer's Disease Neuroimaging Initiative

Jeffery L. Dage<sup>7</sup>, Erik Stomrud<sup>1,2</sup>, Shorena Janelidze<sup>1</sup>, Niklas Mattsson-Carlsson<sup>1,8,9</sup> and Oskar Hansson<sup>1,2</sup>

## Methods

**Participants.** Participants from the Swedish BioFINDER study (<http://biofinder.se>; NCT01208675) consisted of consecutively included non-dementia patients with mild cognitive symptoms who were referred to the participating memory clinics. Two hundred and fifty-three (81%) patients were referred from primary care units; 40 (13%) patients were referred from other specialist settings; and 20 (6%) patients were self-referrals. The inclusion criteria were as follows: (1) referred to the memory clinic at Skåne University Hospital or Ängelholm Hospital in Sweden owing to cognitive symptoms experienced by the patient and/or informant. These symptoms did not have to be memory complaints but could also be executive, visuospatial, language, praxis or psychomotor complaints; (2) age between 60 and 80 years; (3) Mini-Mental State Examination (MMSE) score of 24–30 points at the baseline visit; (4) do not fulfill the criteria for any dementia; and (5) speak and understand Swedish to the extent that an interpreter was not necessary for the patient to fully understand the study information and neuropsychological tests. The exclusion criteria were as follows: (1) significant unstable systemic illness or organ failure, such as terminal cancer, that makes it difficult to participate in the study; (2) current substantial alcohol or substance misuse; (3) refusing lumbar puncture or neuropsychological assessment; and (4) the cognitive impairment at the baseline visit could, with certainty, be explained by another condition or disease, such as normal pressure hydrocephalus, major cerebral hemorrhage, brain infection, brain tumor, multiple sclerosis, epilepsy, psychotic disorders, severe depression and ongoing medication with drugs that invariably cause cognitive impairment (such as high-dose benzodiazepines).

After inclusion, the patients were categorized as having either SCD or MCI based on an extensive neuropsychological battery performed at baseline, examining verbal, episodic memory, visuospatial ability and attention/executive domains, as previously described<sup>35</sup>. Patients with domain *z*-scores of  $\leq -1.5$  in at least one domain were classified as having MCI. In agreement with Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, (DSM-5) criteria for mild neurocognitive disorders, all patients with composite *z*-scores of  $-1$  to  $-1.5$  were individually assessed by a senior neuropsychologist and classified as having MCI if the performance was assessed to represent a significant cognitive decline in comparison with their estimated premorbid level. All patients with SCD were required to experience cognitive symptoms (but not necessarily in the memory domain). Note that the cognitive tests used for the SCD versus MCI differentiation were different from those used as predictors in the models.

Participants with at least one follow-up visit and a complete baseline dataset of all variables included in the logistic regression models were selected for this study. See Extended Data Fig. 1 for an enrollment flowchart. The participants were followed longitudinally with yearly follow-ups that included cognitive testing, informant-based activities of daily living (ADL) questionnaires and detailed assessments by physicians experienced in neurocognitive disorders. All patients gave their written informed consent to participate, and the study was approved by the regional ethics committee in Lund, Sweden.

**The clinical diagnostic prediction.** In a subgroup of patients (those included from the memory clinics in Malmö and Lund), the treating physician at the memory clinic was prospectively registering the most likely underlying cause of the cognitive impairment (here called the ‘clinical prediction’) in the clinical research form at baseline. The clinical prediction was based on the first visit to the clinic (1.5-h-long visit with the patient and informant), informant-based cognitive symptom (CIMP-QUEST<sup>43</sup>) and ADL (FAQ<sup>45</sup>) questionnaires, a broad cognitive test battery (<https://biofinder.se/data-biomarkers/clinical-evaluation/>) and a CT or MRI scan. The physicians had no access to CSF, plasma and PET biomarker data, because these tests were performed after this initial visit.

**Cognitive tests.** Brief cognitive tests that were available in both BioFINDER and ADNI were selected to approximately represent different cognitive domains. Trail-Making Test B and verbal fluency (animals) were selected as measures of executive/attention performance based on their validated use in modified preclinical Alzheimer’s cognitive composites, which are sensitive in tracking cognitive decline in AD<sup>31,33</sup>. The ten-word delayed recall test from the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) has also been validated for detecting early cognitive decline in AD<sup>31,32</sup> and was chosen for the memory domain. The naming objects and fingers task from the ADAS-Cog was used for verbal performance<sup>45</sup>, and the clock-drawing test was used for visuospatial performance<sup>46</sup>. Each domain was converted to a *z*-score based on the test score distribution in the present population. See Supplementary Methods for the *z*-scores used in the cross-validated model. In addition, the MMSE was used as a brief test of global cognition with specific sensitivity to the cognitive decline seen in AD<sup>33,47</sup>.

**Plasma biomarkers.** Blood samples were collected at baseline and analyzed according to a standardized protocol<sup>12</sup>. Plasma P-tau217 and P-tau181 concentrations in BioFINDER were measured on a Meso-Scale Discovery platform, using an assay developed by Eli Lilly, as previously described<sup>14,16,30</sup>. For the plasma P-tau217 assay, biotinylated-IBA493 was used as a capture antibody and SULFO-TAG-4G10-E2 as the detector (both antibodies developed by Lilly Research Laboratory). For plasma P-tau181 assays, biotinylated-AT270 was used as a capture

antibody (Thermo Fisher Scientific, cat. no. MN1050) and SULFO-TAG-LRL (antibody developed by Lilly Research Laboratory) as the detector. Note that plasma P-tau181 was available only in a smaller subsample ( $n=192$ ) and that a different platform, capture antibody and detection antibody were used compared with ADNI (in addition to the differences in pre-analytical protocols), making the measured concentrations incomparable to the plasma P-tau181 concentrations measured in ADNI. Plasma A $\beta$ 42 and A $\beta$ 40 concentrations were analyzed using the Elecsys immunoassays on a cobas e601 analyzer (Roche Diagnostics), and plasma NFL was measured using Simoa, as previously described<sup>10</sup>.

**Plasma P-tau in the cross-validated model.** Cutoffs were developed to evaluate the generalizability of plasma P-tau-based models across cohorts and assays. The cutoffs were established using an out-of-sample population to further add to the robustness of the results. In our previous work in neuropathology-confirmed cases<sup>14,16</sup>, we observed that changes in plasma P-tau are associated with tau pathology only in the presence of A $\beta$  pathology. This was in contrast to cases with low A $\beta$  where very few cases had elevated plasma P-tau. Furthermore, we found that fewer cognitively normal individuals had increased plasma P-tau levels in contrast to those with MCI and AD dementia. Based on these previous findings, we used A $\beta$ -negative cognitively normal individuals to define an independent normal population and establish cutoffs for plasma P-tau217 and P-tau181 in BioFINDER and ADNI, respectively. In BioFINDER, CSF A $\beta$ 42/A $\beta$ 40 was used to define A $\beta$  status, and the A $\beta$ -negative sample consisted of 215 healthy controls. In ADNI, <sup>18</sup>F-florbetapir PET was used to define A $\beta$  status and the A $\beta$ -negative sample consisting of 547 healthy controls. The cutoff for plasma P-tau abnormality was set at the mean + 2 s.d. in each sample, respectively, similarly to how plasma P-tau cutoffs have been established previously<sup>16</sup>. This approach, thus, resulted in unbiased, non-optimized cutoffs, which should provide higher reproducibility in other populations<sup>48</sup>. The cutoff was  $>0.387$  pg ml<sup>-1</sup> for plasma P-tau217 and  $>38.2$  pg ml<sup>-1</sup> for plasma P-tau181 (note that differences in P-tau isoforms, assays and platforms explain the large difference in measured concentrations).

In a secondary cross-validated model, *z*-scored plasma P-tau data were used. The *z*-scores for plasma P-tau217 and P-tau181 were established based on the distribution in the above-described A $\beta$ -negative samples the following way: (P-tau concentration – mean P-tau concentration in reference sample) / s.d. of P-tau concentrations in reference sample. Such *z*-scores can, thus, be obtained from any clinical chemistry lab with a similar reference sample.

**CSF biomarkers.** CSF was collected and handled according to a structured protocol as previously described<sup>1</sup>. P-tau and the A $\beta$ 42/A $\beta$ 42 were analyzed using the Elecsys immunoassays on a cobas e601 analyzer<sup>49</sup>. NFL was analyzed as previously described<sup>1</sup>.

**MRI.** The MRI protocol for BioFINDER was previously described<sup>50</sup>. As MRI measure, cortical thickness in temporal brain regions susceptible to atrophy in AD was used (referred to as the ‘AD signature’ region). The cortical thickness was quantified using FreeSurfer version 5.1 (<http://surfer.nmr.mgh.harvard.edu>). The AD signature region was calculated based on the cortical thickness in entorhinal, inferior temporal, middle temporal and fusiform regions, as previously described<sup>18</sup>. The AD signature region was chosen instead of hippocampal volume, which performed poorer for predicting progression to AD dementia (data not shown).

**Outcomes.** The primary outcome was prediction of progression to AD dementia versus progression to any other dementia or not progressing to any dementia within 4 years. Four years was chosen to reflect a clinically relevant timeframe in which it seems reasonable for a physician to give prognostic advice to an elderly patient and also a suitable timeframe for clinical trials to detect differences in conversion to dementia. Those who converted to AD dementia within that timeframe were coded as ‘1’, and stable SCD/MCI and conversion to any other dementia within the timeframe were coded as ‘0’. Non-dementia converters with follow-ups <4 years were excluded from this analysis. Conversion to AD dementia within 2 and 6 years, respectively, were secondary outcomes with corresponding selection criteria for the examined population. The follow-up diagnosis was based on the treating physician’s follow-up assessments and reviewed by a consensus group including memory clinic physicians and a senior neuropsychologist. The diagnosis was based on the DSM-5 criteria for major neurocognitive disorder due to probable AD. In addition, the patient was required to show signs of abnormal amyloid accumulation according either to CSF analysis<sup>49</sup> or A $\beta$  PET<sup>1</sup> in agreement with the National Institute on Aging–Alzheimer’s Association criteria for AD<sup>51</sup>. See Supplementary Methods for further details. Plasma biomarkers were analyzed after the follow-up diagnosis had been determined (that is, neither the treating physician nor the consensus group had access to them).

**Validation cohort.** The ADNI was used as the validation cohort (NCT00106899). The data used in the preparation of this article were obtained from the ADNI database (<http://adni.loni.usc.edu>). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers and clinical and neuropsychological assessment can be combined to

measure the progression of MCI and early AD. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org). ADNI was approved by the institutional review boards of all participating institutions, and written informed consent was obtained from all participants at each site (see the Reporting Summary for a list of the institutional review boards).

According to our aim, we selected only non-dementia patients with cognitive symptoms at baseline. This included participants with MCI from the MCI cohort and participants from the healthy control cohort who had significant memory concerns (here referred to as 'SCD'). Inclusion/exclusion criteria are described in detail at [www.adni-info.org](http://www.adni-info.org) and in Supplementary Methods. Briefly, all patients in this study were between the ages of 55 and 91 years, had completed at least 6 years of education, were fluent in Spanish or English and were free of any significant neurologic disease other than AD. Patients with SCD had an MMSE score  $\geq 24$  and a Clinical Dementia Rating (CDR) score of 0 but expressed concerns about memory impairment on the Cognitive Change Index<sup>32</sup>. Patients classified as having MCI had an MMSE score  $\geq 24$ , objective memory loss as shown on scores on delayed recall on the Wechsler Memory Scale Logical Memory II test, a CDR of 0.5, preserved ADL and absence of dementia. Detailed criteria for SCD and MCI are specified in the Supplementary Methods. All SCD and MCI participants with at least one follow-up visit and a complete dataset of variables included in the logistic regression models were included in this study. The variables used in the model selection were the same as in BioFINDER except for the plasma P-tau biomarker (in ADNI, P-tau181 was available instead of P-tau217).

The plasma-handling procedures were described previously<sup>27</sup>. Plasma P-tau181 in ADNI was measured on Simoa HD-X instruments (Quanterix) at the Clinical Neurochemistry Laboratory, University of Gothenburg. The capture antibody AT270 (MN1050, Invitrogen) was coated onto paramagnetic beads (103207, Quanterix), and the detector antibody (Tau12, 806502, BioLegend) was biotinylated; these reagents were used together with recombinant tau 441 phosphorylated in vitro by glycogen synthase kinase 3 $\beta$  (TO8-50FN, SignalChem) as the calibrator to build the assay, as described previously<sup>45</sup>. Personnel performing the plasma analysis were blinded to the clinical and biomarker information. The analysis was performed after the follow-up diagnosis had been determined. Further assay details were published previously<sup>27</sup>. Plasma NfL was analyzed using the same Simoa-based assay as described for the BioFINDER study. In ADNI, there were two different plasma A $\beta$ 42/A $\beta$ 40 assays, and each one was available only in very small datasets. Therefore, plasma A $\beta$ 42/A $\beta$ 40 was not included in the ADNI models. This is further described in the Supplementary Methods. CSF P-tau181 was analyzed using an Elecsys immunoassay on a cobas e601 analyzer, as previously described<sup>46</sup>.

The MRI measure was extracted from structural brain images acquired using 3T MRI scanners with T1-weighted scans. Cortical thickness regions were quantified using FreeSurfer version 5.1 (<http://surfer.nmr.mgh.harvard.edu>) and combined to the AD signature composite region, as described in the BioFINDER methods.

To create a similar outcome variable as in BioFINDER, participants were deemed converters to AD dementia if they had a follow-up diagnosis of AD dementia<sup>53</sup> and were A $\beta$ -positive according to the A $\beta$  PET scan<sup>54</sup>. Cognitively stable participants and converters to other dementias or A $\beta$ -negative AD dementia were, thus, coded as non-AD dementia converters. See Supplementary Methods for further details.

**Statistical analysis.** Conversion to AD dementia was used as the dependent variable in logistic regression models. All continuous independent variables were transformed to z-scores based on the distribution in the present population. *APOE*  $\epsilon 4$  genotype was coded into two different variables: presence of just one  $\epsilon 4$  allele and presence of two  $\epsilon 4$  alleles, as per previously described differences in their risk of AD<sup>32,34</sup>. The initial model selection was performed using the R package MuMIn, which tests all different combinations of variables and then ranks the models according to the AIC. For more information on the AIC, see Supplementary Methods and refs. 19,20. The model with the lowest AIC was selected as the model with the best tradeoff between fit and model complexity. The next step was then to find models with as few variables as possible that had a similar performance (defined as  $\Delta AIC < 2$  from the model with the lowest AIC<sup>19,20,30</sup>). Therefore, a stepwise removal of variables was performed as long as the  $\Delta AIC$  was  $< 2$  from the model with best fit to end up with the 'parsimonious model'. In addition, only variables with  $P < 0.10$  were kept in the parsimonious model. Then, variables were removed using a stepwise procedure in subsequent models to illustrate the added value of the different variables to plasma P-tau. This process was repeated for the time points 2, 4 and 6 years. Comparisons of AUC were performed using DeLong statistics. A two-sided  $P$  value  $< 0.05$  was considered statistically significant. R version 4.0 was used for all statistical analyses.

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

## Data availability

For BioFINDER data, 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 and as long as data transfer is in agreement with EU legislation on the general data protection regulation and decisions by the Ethical Review Board of Sweden and Region Skåne, which should be regulated in a material transfer agreement. ADNI data are stored (publicly available) in the loni database (<https://ida.loni.usc.edu/>).

## Code availability

No custom code or mathematical algorithm that was central to the conclusions was used in this study.

## References

- Åstrand, R., Rolstad, S. & Wallin, A. Cognitive Impairment Questionnaire (CIMP-QUEST): reported topographic symptoms in MCI and dementia. *Acta Neurol. Scand.* **121**, 384–391 (2009).
- Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H. Jr., Chance, J. M. & Filos, S. Measurement of functional activities in older adults in the community. *J. Gerontol.* **37**, 323–329 (1982).
- Rosen, W. G., Mohs, R. C. & Davis, K. L. A new rating scale for Alzheimer's disease. *Am. J. Psychiatry* **141**, 1356–1364 (1984).
- Shulman, K. I. Clock-drawing: is it the ideal cognitive screening test? *Int J. Geriatr. Psychiatry* **15**, 548–561 (2000).
- Palmqvist, S. et al. Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. *PLoS ONE* **7**, e38639 (2012).
- Mattsson-Carlgen, N., Palmqvist, S., Blennow, K. & Hansson, O. Increasing the reproducibility of fluid biomarker studies in neurodegenerative studies. *Nat. Commun.* **11**, 6252 (2020).
- Hansson, O. et al. CSF biomarkers of Alzheimer's disease concord with amyloid- $\beta$  PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement.* **14**, 1470–1481 (2018).
- Palmqvist, S. et al. Earliest accumulation of  $\beta$ -amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nat. Commun.* **8**, 1214 (2017).
- 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.* **7**, 263–269 (2011).
- Aisen, P. S., Petersen, R. C., Donohue, M. & Weiner, M. W. & Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Neuroimaging Initiative 2 Clinical Core: progress and plans. *Alzheimers Dement.* **11**, 734–739 (2015).
- McKhann, G. et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–944 (1984).
- Landau, S. M. et al. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann. Neurol.* **72**, 578–586 (2012).

## Acknowledgements

Work at the authors' research center was supported by the Swedish Research Council (2016-00906 (O.H.) and 2018-02052 (S.P.)), the Knut and Alice Wallenberg Foundation (2017-0383 (O.H.)), the Marianne and Marcus Wallenberg Foundation (2015.0125 (O.H.)), the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation (AF-745911 (O.H.) and AF-940046 (S.P.)), the Swedish Brain Foundation (FO2019-0326 (O.H.) and FO2020-0271 (S.P.)), the Parkinson Foundation of Sweden (1280/20 (O.H.)), the Skåne University Hospital Foundation (2020-0000028 (S.P.)), Regionalt Forskningsstöd (2020-0314 (O.H.) and 2020-0383 (S.P.)) and the Swedish federal government under the ALF agreement (2018-Projekt0279 (O.H.) and 2018-Projekt0226 (S.P.)). H.Z. is a Wallenberg Scholar supported by grants from the Swedish Research Council (2018-02532), the European Research Council (681712), Swedish State Support for Clinical Research (ALFGBG-720931), the Alzheimer Drug Discovery Foundation, USA (201809-2016862), AD Strategic Fund and the Alzheimer's Association (ADSF-21-831376-C, ADSF-21-831381-C and ADSF-21-831377-C), the European Union's Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie grant agreement no. 860197 (MIRIADE) and the UK Dementia Research Institute at University College London. K.B. is supported by the Swedish Research Council (2017-00915), the Alzheimer Drug Discovery Foundation, USA (RDAPB-201809-2016615), the Swedish Alzheimer Foundation (AF-742881), Hjärnfonden, Sweden (FO2017-0243), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF agreement (ALFGBG-715986) and the European Union Joint Program for Neurodegenerative Disorders (JPND2019-466-236). The precursor of <sup>18</sup>F-flutemetamol was sponsored by GE Healthcare. The precursor of <sup>18</sup>F-RO948 was provided by Roche. ADNI data collection and sharing was funded by the Alzheimer's Disease Neuroimaging Initiative (National Institutes of Health grant U01 AG024904) and Department of Defense award no. W81XWH-12-2-0012. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering and through

generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica; Biogen; Bristol-Myers Squibb Company; CereSpir; Cogstate; Eisai; Elan Pharmaceuticals; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche and its affiliated company, Genentech; Fujirebio; GE Healthcare; IXICO; Janssen Alzheimer Immunotherapy Research & Development; Johnson & Johnson Pharmaceutical Research & Development; Luminosity; Lundbeck; Merck & Co.; Meso Scale Diagnostics; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals; Pfizer; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. ADNI investigators contributed to the design and implementation of the ADNI database and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf). Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

### Author contributions

S.P., P.T., N.C., H.Z., K.B., J.L.D., E.S., S.J., N.M.-C. and O.H. collected the data and reviewed the manuscript for intellectual content. S.P. drafted the manuscript, analyzed the data and prepared the figures. S.P. and O.H. interpreted the data, and N.C. provided statistical support. N.M.-C. implemented the models online. O.H. was the principal investigator of BioFINDER and supervised collection, analysis and interpretation of the study data.

### Competing interests

S.P. has served on the scientific advisory boards for Hoffman-La Roche and Geras Solutions. H.Z. has served on the scientific advisory boards for Denali, Roche Diagnostics, Wave, Samumed, Siemens Healthineers, Pinteon Therapeutics and CogRx, has given lectures in symposia sponsored by Fujirebio, Alzecure and Biogen and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. K.B. has served as a consultant, on advisory boards or on the data monitoring committees for Abcam, Axon, Biogen, JOMDD/Shimadzu, Julius Clinical, Eli Lilly, MagQu, Novartis, 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. J.L.D. is an employee of Eli Lilly and Company. O.H. has acquired research support (for the institution) from AVID Radiopharmaceuticals, Biogen, Eli Lilly, Eisai, GE Healthcare, Pfizer and Roche. In the past 2 years, O.H. has also received consultancy/speaker fees from AC Immune, Alzpath, Biogen, Cerveau and Roche. S.J., P.T., N.C., E.S. and N.M.-C. report no disclosures.

### Additional information

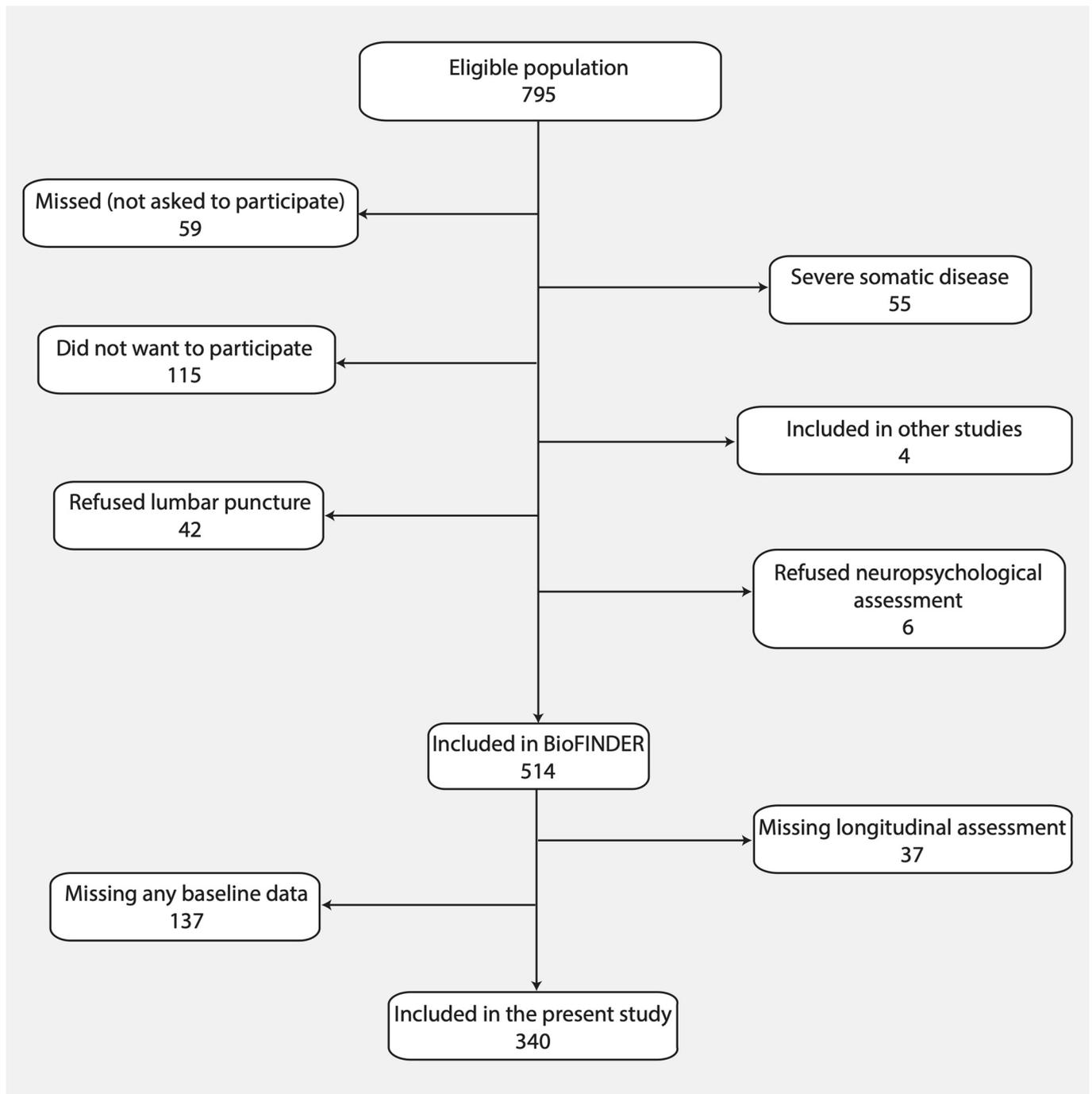
**Extended data** is available for this paper at <https://doi.org/10.1038/s41591-021-01348-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41591-021-01348-z>.

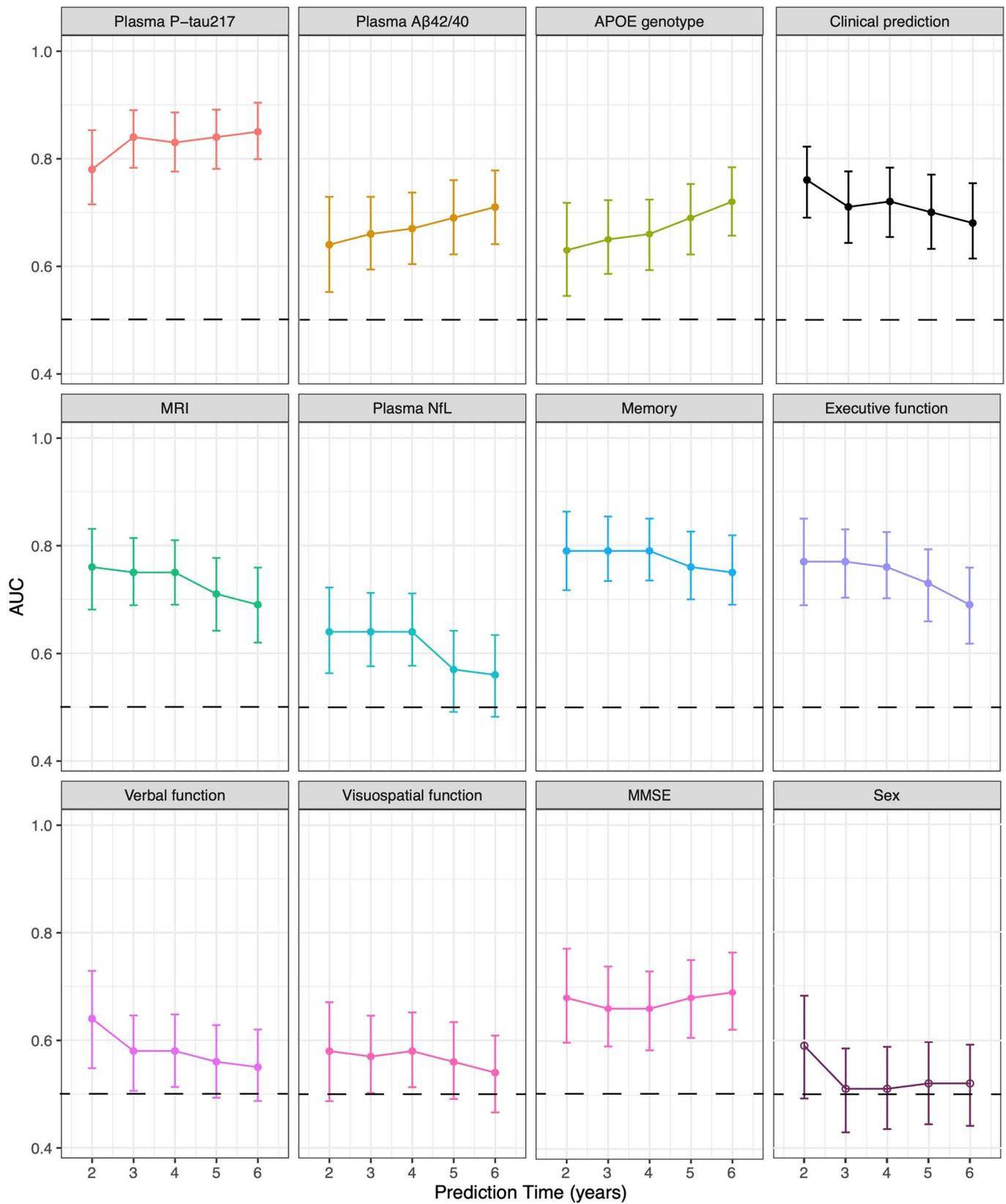
**Correspondence and requests for materials** should be addressed to S.P. or O.H.

**Peer review information** *Nature Medicine* thanks Ronald Petersen, Stephen Salloway and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Jerome Staal was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

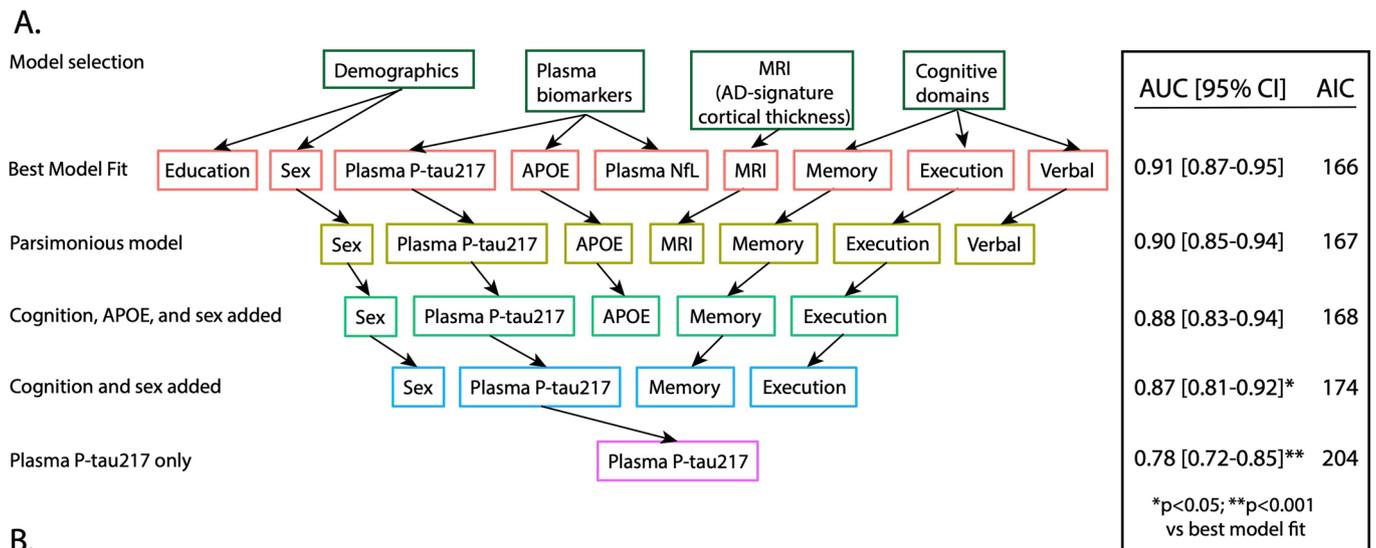


**Extended Data Fig. 1 | Enrollment flowchart for the BioFINDER sample.** Eligible population was defined as being referred to any of the participating memory clinics and being non-demented.

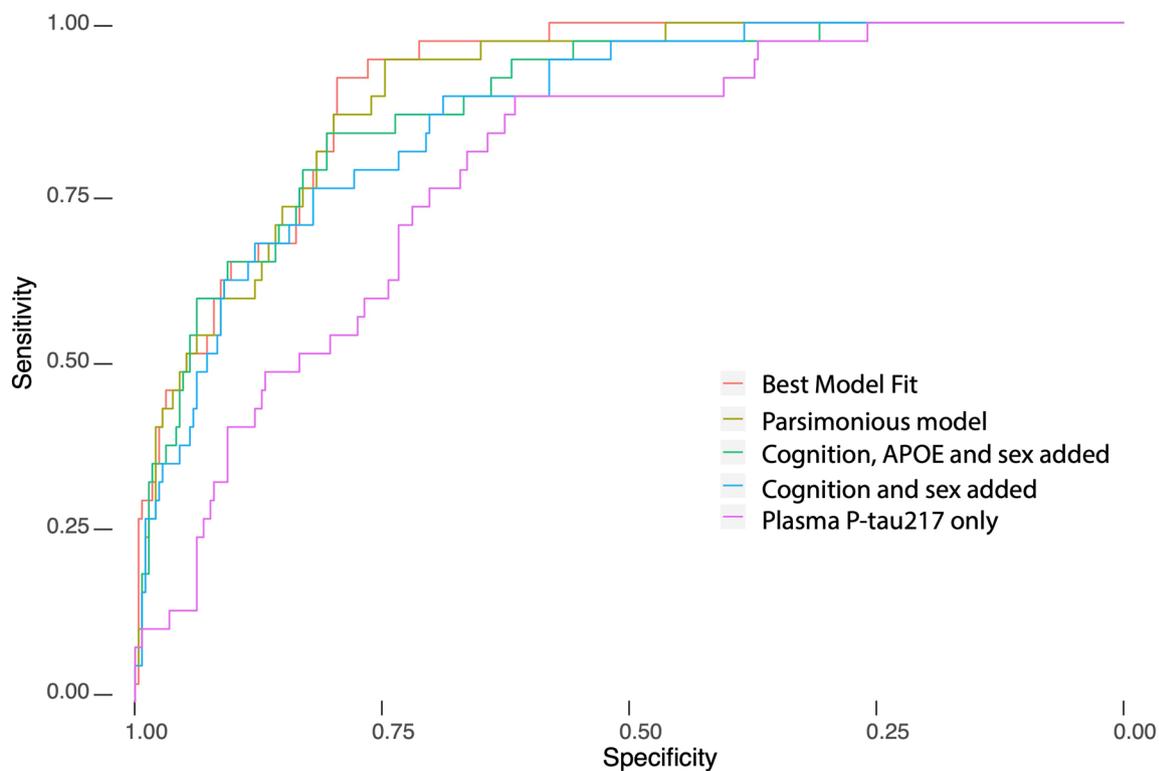


Extended Data Fig. 2 | See next page for caption.

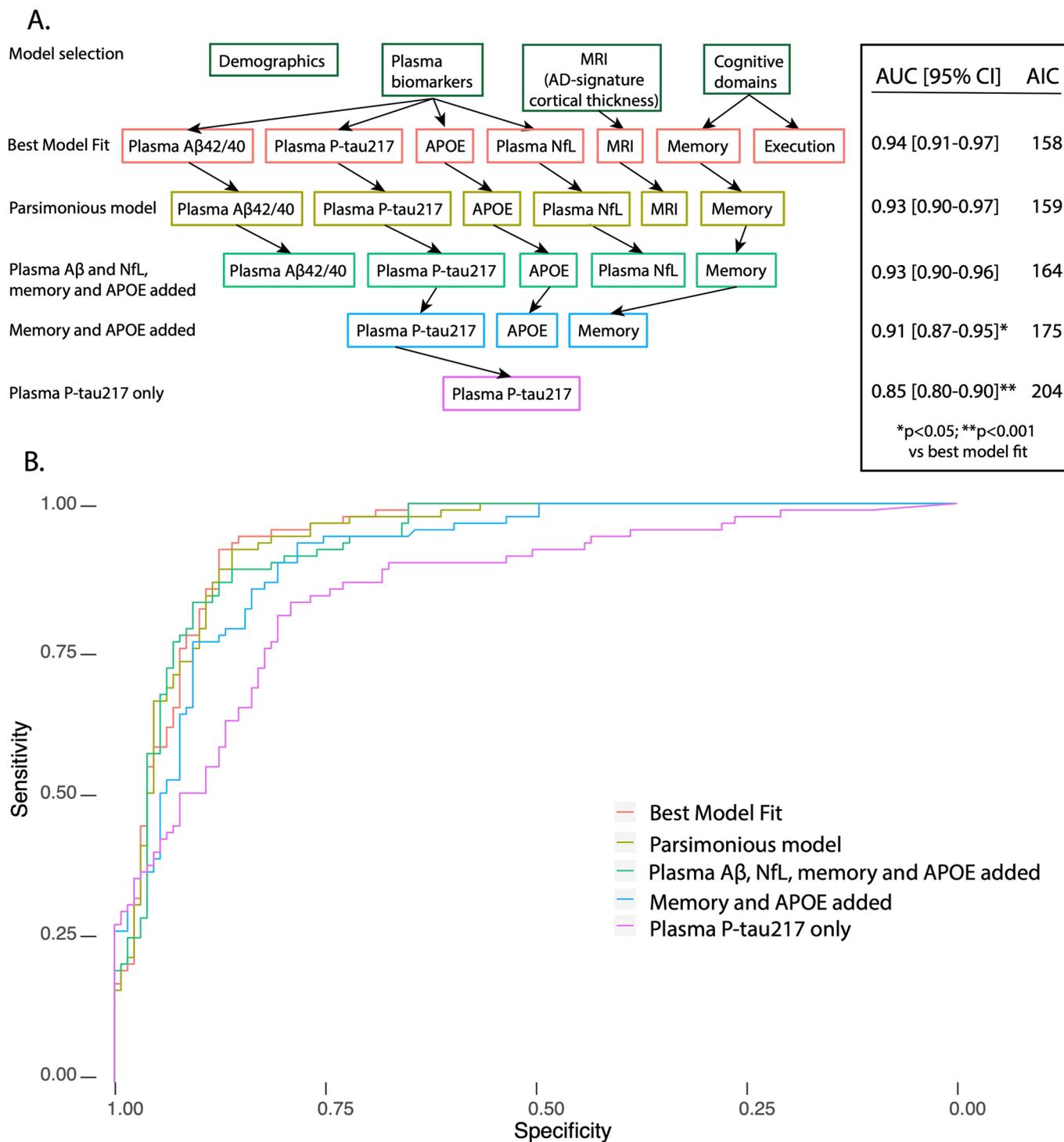
**Extended Data Fig. 2 | Univariable logistic regression models for predicting progression to AD dementia vs other conditions within 2–6 years.** Data are shown as AUCs at each time point (error bars show the 95% CIs of the AUCs). AUCs above the dashed lines represent a predictive accuracy better than chance (AUC 0.5). *APOE* genotype was coded as 0, 1 or 2  $\epsilon 4$  alleles. Regarding the cognitive measures, memory had high accuracy from short- to long-term predictions, while executive function had lower accuracies for long-term prediction. This suggests that memory changes earlier than executive function during the development of AD. Regarding the biomarkers, cortical thickness representing AD-specific neurodegeneration were best for short- to mid-term prediction, plasma P-tau217 for mid- to long-term prediction and plasma A $\beta$ 42/A $\beta$ 40 better for long-term prediction. This is congruent with the model for the development of AD that begins with the accumulation of A $\beta$ , then phosphorylation of tau and the deposition of tau tangles, and finally neurodegeneration.



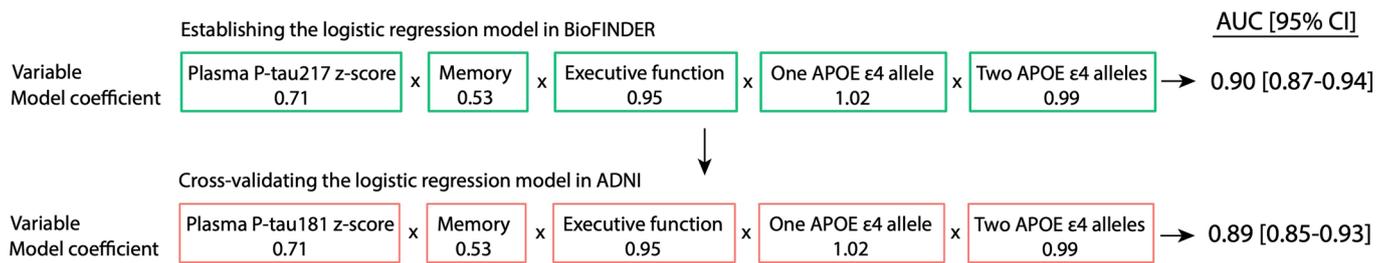
**B.**



**Extended Data Fig. 3 | Model selection process and performance for predicting AD dementia within 2 years in BioFINDER. a,** Model selection process. Best Model Fit shows the data-driven model selection with the lowest AIC (that is, the best model fit). The parsimonious model shows the model that had a similar model fit ( $\Delta AIC < 2$ ) with as few significant predictors as possible. In subsequent models, the least important modalities were removed in a step-wise procedure. Model specifications including comparisons between all models are shown in Supplementary Table 2. **b,** ROC curves of the different models. Abbreviations: AD, Alzheimer's disease; AIC, Akaike Information Criterion; *APOE*, Apolipoprotein E genotype (number of  $\epsilon 4$  alleles); AUC, Area under the ROC curve; MRI, Cortical thickness of a temporal AD-specific region; ROC, Receiver Operating Characteristic.



**Extended Data Fig. 4 | Model selection process and performance for predicting AD dementia within 6 years in BioFINDER. a,** Model selection process. Best Model Fit shows the data-driven model selection with the lowest AIC (*that is*, the best model fit). The parsimonious model shows the model that had a similar model fit ( $\Delta AIC < 2$ ) with as few significant predictors as possible. In subsequent models, the least important modalities were removed in a step-wise procedure. Model specifications including comparisons between all models are shown in Supplementary Table 3. **b,** ROC curves of the different models. Abbreviations: AD, Alzheimer’s disease; AIC, Akaike Information Criterion; APOE, Apolipoprotein E genotype (number of  $\epsilon 4$  alleles); AUC, Area under the ROC curve; MRI, Cortical thickness of a temporal AD-specific region; ROC, Receiver Operating Characteristic.



**Extended Data Fig. 5 | Cross-validation and implementation of an algorithm using plasma P-tau z-scores.** Plasma P-tau z-scores based on the distribution of Aβ-negative cognitively unimpaired participants in BioFINDER and ADNI, respectively (see Methods), was used in the logistic regression models. Model coefficients were established in BioFINDER (AUC 0.90) and tested in ADNI (AUC 0.89). Cognitive z-scores have been inverted so that higher scores equal poorer results. This model is implemented at <http://predictAD.app> where one can enter the raw cognitive test scores that constitute the z-scores, number of APOE ε4 alleles and plasma P-tau z-score (either from P-tau217 or P-tau181).

## Reporting Summary

Nature Research 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 Research 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** For ADNI data: Data is exported to csv-files and stored (publicly available) at the Ioni database. For BioFINDER data: Data is exported to csv-files and stored at Skåne University Hospital, Sweden.

**Data analysis** R programming language was used (ver 4.0) and Freesurfer (ver 5.1)

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 Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

For BioFINDER data: 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 and 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. For ADNI data: Data is stored (publicly available) at the Ioni database (<https://ida.ioni.usc.edu/>).

## 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](https://www.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	The study included two prospective studies (BioFINDER [n=340] and ADNI [n=543]) with large sample sizes. All participants with SCD or MCI at baseline with available data were analyzed in this study.
Data exclusions	Only those with missing data for any of the variables in any models were excluded. See flowchart in Supplementary Fig. 1.
Replication	A cross-validated model is presented. The logistic regression model was established in BioFINDER and the model was replicated in the independent ADNI cohort. In addition to that, data-driven variable selections for logistic regression models was performed separately in the two independent cohorts with similar results.
Randomization	In these two prospective cohort studies (observational studies) no allocation into experimental groups were performed, therefore randomization is not relevant to this study. Statistical analyses were controlled for potential confounding effect of age and sex.
Blinding	All plasma and CSF analyses were performed by individuals who were blinded to the clinical data. The clinical diagnostic prediction by memory clinic physicians was performed blinded to plasma and CSF data (i.e. performed before plasma and CSF was analysed).

## 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	Involved 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"/> Human research participants
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved 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 P-tau217 and P-tau181 concentrations in BioFINDER were measured on a Meso-Scale Discovery platform (MSD, Rockville, MD), using an assay developed by Eli Lilly as previously described.[14,16,30] For the plasma P-tau217 assay, biotinylated-IBA493 was used as a capture antibody and SULFO-TAG-4G10-E2 as the detector (both antibodies developed by Lilly Research Laboratory). For plasma P-tau181 assays, biotinylated-AT270 was used as a capture antibody (Thermo Fisher, catalog number: MN1050) and SULFO-TAG-LRL (antibody developed by Lilly Research Laboratory) as the detector. Plasma P-tau181 in ADNI was measured on Simoa HD-X instruments (Quanterix, Billerica, MA, USA) at the Clinical Neurochemistry Laboratory, University of Gothenburg, Mölndal, Sweden, as previously described.[27] The capture antibody AT270 (#MN1050, Invitrogen) was coated onto paramagnetic beads (#103207, Quanterix) and the detector antibody (Tau12, #806502, BioLegend) was biotinylated; these reagents were used together with recombinant tau 441 phosphorylated in vitro by glycogen synthase kinase 3 $\beta$ (#TO8-50FN, SignalChem) as the calibrator to build the assay, as described before. [15]
Validation	This has been published previously in several paper, for example by Palmqvist et al. JAMA 2020 (the papers are referred to in the manuscript).

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Detailed information is given in Table 1 and in the Methods section. Briefly, BioFINDER included 164 SCD and 176 MCI participants of whom 91 progressed to AD dementia at follow-up, 49 to other dementias and 200 who did not progress to
----------------------------	--

any dementia. The mean (standard deviation [SD]) age was 70.7 (5.6) years and 49% were women. The mean (SD) time to dementia was 2.9 (1.5) years and the mean (SD) follow-up time in participants that did not progress to dementia was 4.5 (1.6) years. The validation cohort (ADNI) included 106 SCD and 437 MCI participants of whom 102 progressed to AD dementia at follow-up and 28 to other dementias. In the main cohort (BioFINDER) all participants were referred from physicians (mostly general practitioners) as part of clinical routine practice and then consecutively included in the study. This should eliminate any self-selection bias or other biases. The SCD participants in the validation cohort (ADNI) were recruited through advertisement and the MCI participant were not consecutively recruited, which do not rule out potential recruitment biases (discussed in the manuscript).

#### Recruitment

In BioFINDER, the sample consisted of patients that had been referred to participating memory clinics (mostly from primary care). The patients were recruited consecutively between 2010 and 2014 (see Supplementary Fig. 1 for an enrollment flowchart). In the ADNI study, participants were recruited non-consecutively from many different centres between 2005 and 2019, and thereby represent a more selected sample (i.e. closer to a clinical trial population). Despite this difference, our findings held when validated in both cohorts.

#### Ethics oversight

For BioFINDER, all patients gave their written informed consent to participate, and the study was approved by the regional ethics committee in Lund, Sweden. ADNI was approved by the institutional review boards of all participating institutions and written informed consent was obtained from all participants at each site. The Ethics committees/institutional review boards that approved the ADNI study are:

Albany Medical Center Committee on Research Involving Human Subjects Institutional Review Board, Boston University Medical Campus and Boston Medical Center Institutional Review Board, Butler Hospital Institutional Review Board, Cleveland Clinic Institutional Review Board, Columbia University Medical Center Institutional Review Board, Duke University Health System Institutional Review Board, Emory Institutional Review Board, Georgetown University Institutional Review Board, Health Sciences Institutional Review Board, Houston Methodist Institutional Review Board, Howard University Office of Regulatory Research Compliance, Icahn School of Medicine at Mount Sinai Program for the Protection of Human Subjects, Indiana University Institutional Review Board, Institutional Review Board of Baylor College of Medicine, Jewish General Hospital Research Ethics Board, Johns Hopkins Medicine Institutional Review Board, Lifespan - Rhode Island Hospital Institutional Review Board, Mayo Clinic Institutional Review Board, Mount Sinai Medical Center Institutional Review Board, Nathan Kline Institute for Psychiatric Research & Rockland Psychiatric Center Institutional Review Board, New York University Langone Medical Center School of Medicine Institutional Review Board, Northwestern University Institutional Review Board, Oregon Health and Science University Institutional Review Board, Partners Human Research Committee Research Ethics, Board Sunnybrook Health Sciences Centre, Roper St. Francis Healthcare Institutional Review Board, Rush University Medical Center Institutional Review Board, St. Joseph's Phoenix Institutional Review Board, Stanford Institutional Review Board, The Ohio State University Institutional Review Board, University Hospitals Cleveland Medical Center Institutional Review Board, University of Alabama Office of the IRB, University of British Columbia Research Ethics Board, University of California Davis Institutional Review Board Administration, University of California Los Angeles Office of the Human Research Protection Program, University of California San Diego Human Research Protections Program, University of California San Francisco Human Research Protection Program, University of Iowa Institutional Review Board, University of Kansas Medical Center Human Subjects Committee, University of Kentucky Medical Institutional Review Board, University of Michigan Medical School Institutional Review Board, University of Pennsylvania Institutional Review Board, University of Pittsburgh Institutional Review Board, University of Rochester Research Subjects Review Board, University of South Florida Institutional Review Board, University of Southern California Institutional Review Board, UT Southwestern Institutional Review Board, VA Long Beach Healthcare System Institutional Review Board, Vanderbilt University Medical Center Institutional Review Board, Wake Forest School of Medicine Institutional Review Board, Washington University School of Medicine Institutional Review Board, Western University Institutional Review Board, Western University Health Sciences Research Ethics Board, and Yale University Institutional Review Board.

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

## 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

NCT01208675 (BioFINDER) and NCT00106899 (ADNI).

#### Study protocol

For BioFINDER: [www.biofinder.se](http://www.biofinder.se). For ADNI: [www.adni-info.org](http://www.adni-info.org) or [ida.loni.usc.edu](http://ida.loni.usc.edu)

#### Data collection

2010 - 2020 for BioFINDER and 2005 - 2019 for ADNI. BioFINDER data are collected at the memory clinics of Skåne University Hospital and Ängelholm's hospital in Sweden. ADNI data are collected at 61 sites in the US and Canada as described at <http://adni.loni.usc.edu/>

#### Outcomes

The primary outcome was prediction of progression to AD dementia vs progression to any other dementia or not progressing to any dementia within 4 years. Four years was chosen to reflect a clinically relevant timeframe in which it seems reasonable for a physician to give prognostic advice to an elderly patient and also a suitable timeframe for clinical trials to detect differences in conversion to dementia. Conversion to AD dementia within 2 and 6 years, respectively, were secondary outcomes. The follow-up diagnosis was based on the treating physician's follow-up assessments and reviewed by a consensus group including memory clinic physicians and a senior neuropsychologist. The diagnosis was based on the DSM-5 criteria for major neurocognitive disorder due to probable AD. In addition, the patient was required to show signs of abnormal amyloid accumulation either according to CSF analysis or A $\beta$  PET.