CORTEX 202 (2026) 181–192

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

A multiverse analysis of the logical memory test and plasma biomarkers of Alzheimer's disease

Davide Bruno [a,b,*,1], Ainara Jauregi-Zinkunegi [a,1], Rachel L. Studer [c,d], Rachael Wilson [d,e], Henrik Zetterberg [e,f,g,h,i,j,k,l], Sterling C. Johnson [c,d,e] and Kimberly D. Mueller [b,e]

[a] School of Psychology, Liverpool John Moores University, Liverpool, UK
[b] Department of Communication Sciences and Disorders, University of Wisconsin–Madison, Wisconsin, USA
[c] Wisconsin Alzheimer's Institute, School of Medicine and Public Health, University of Wisconsin–Madison, Wisconsin, USA
[d] Department of Medicine, University of Wisconsin–Madison, Wisconsin, USA
[e] Wisconsin Alzheimer's Disease Research Center, University of Wisconsin–Madison, Wisconsin, USA
[f] Department of Pathology and Laboratory Medicine, University of Wisconsin–Madison, Wisconsin, USA
[g] Department of Psychiatry and Neurochemistry, Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden
[h] Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden
[i] Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK
[j] UK Dementia Research Institute at UCL, London, UK
[k] Hong Kong Center for Neurodegenerative Diseases, InnoHK, Hong Kong, China
[l] Centre for Brain Research, Indian Institute of Science, Bangalore, Karnataka, India

ARTICLE INFO

Article history:
Received 10 November 2025
Revised 1 May 2026
Accepted 3 June 2026
Action editor Gail Robinson
Published online 16 June 2026

Keywords:
Logical memory test
Plasma biomarkers
p-tau217
Alzheimer's disease
Sex differences

ABSTRACT

Background: Alzheimer's disease (AD) detection with blood-based biomarkers promises to revolutionise dementia diagnosis. However, blood testing is still a challenge in remote settings. Cognitive testing that is sensitive to plasma biomarker levels can be a useful proxy. A common test for dementia assessment is the logical memory test (LMT) – where a story is read out loud to the individual and then freely recalled. Alongside standard metrics, item-based metrics, such as recall of proper names and serial positions, are effective at detecting AD-related pathology. We set out to compare a range of LMT metrics against plasma AD biomarkers, and examine differences between men and women.

Methods: The Wisconsin Registry for Alzheimer's Prevention cohort was used for analysis: participants (n = 1195, 69% women; mean age = 67.2, SD = 7.7) were free from dementia. Data included logical memory performance, demographics, clinical and genetic information, and plasma biomarkers. Analyses were cross-sectional with a maximum of two years between assessment and biomarker extraction. We carried out multiverse analyses to allow comparison of alternative models with permutations of covariates.

Results: In the full sample, associations were most stable between LMT scores and p-tau217, and were significant across all models, while associations with other plasma biomarkers

* *Corresponding author.*
E-mail address: d.bruno@ljmu.ac.uk (D. Bruno).
[1] contributed equally.

https://doi.org/10.1016/j.cortex.2026.06.005

were generally less reliable. In the stratified analyses, associations were overall more robust and consistent in women than in men.

Discussion: Current findings show LMT metrics are consistently associated with *in vivo* levels of pathology as measured by plasma p-tau217. The discrepancies observed between men and women require further investigation.

1. Background

Alzheimer's disease (AD) detection with blood-based biomarkers promises to revolutionise dementia diagnosis. Plasma phosphorylated tau 217 (p-tau217) has been shown to predict AD with considerable accuracy (Palmqvist et al., 2025; Xiao et al., 2024; Yakoub et al., 2025), and to be AD-specific (VandeVrede et al., 2025). Positive results have also been obtained outside of typical North American and European cohorts (Pandey et al., 2025; Xiao et al., 2024; Zhong et al., 2025). Nevertheless, blood testing for AD still presents a few challenges (Balogun et al., 2023), such as: failing to detect very low levels, which is critical when assessing individuals free of dementia (Karikari et al., 2022); cost (Karikari, 2022); and diurnal variation (Della Monica et al., 2024). Additionally, blood testing in regions where access to fully equipped laboratories is limited (Ansu-Mensah et al., 2024) or in prison settings (Bellass et al., 2024) is challenging.

Neuropsychological assessment is non-invasive, relatively inexpensive, and utilizes accessible technologies (Jauregi-Zinkunegi et al., 2025a), consistent with a low-tech approach (Sarfati et al., 2024). Cognitive testing that is sensitive to plasma biomarkers, therefore, can be a useful proxy if necessary (Bruno et al., 2023). A common test for dementia assessment is story recall (Baek et al., 2011), and the logical memory test (LMT (Wechsler, , 1987);), in particular, has been used to define mild cognitive impairment (MCI) in Alzheimer's Disease Neuroimaging Initiative (ADNI) studies ((Khachaturian et al., 2025); but see (Duff et al., 2021) for a critique). LMT provides consistent results compared to other story recall tests, such as the Craft Story 21 test (Bruno et al., 2024; Nester et al., 2023), and has shown concordance with AD biomarker outputs in cerebrospinal fluid (CSF (Bruno et al., 2023), (Hale et al., 2024)) and positron emission tomography (PET (Terao et al., 2025), (Bruno et al., 2021)). However, it is still possible to extract a richer cognitive signal from LMT without the need to change the task itself, as done previously with other tests (e.g. (Duff et al., 2007),).

Alongside standard scores, in fact, a few process-based metrics have been applied to the LMT, showing effectiveness at detecting AD-related pathology. For example, in a previous study (Jauregi-Zinkunegi et al., 2025b), LMT metrics were also successfully associated with plasma p-tau217 levels. In this research, prediction of p-tau217 was improved when measuring how many idea units were forgotten between immediate and delayed recall. This measure of story forgetting (dubbed total ratio, Tr) is a process score, meaning that it attempts to quantify a cognitive process that underlies test scores (Libon et al., 2022; Rozenblatt, 2018). Moreover, individuals with AD struggle to retrieve proper names significantly more than age-matched controls (Semenza et al., 2003) and this may be owed to degradation of both phonological and semantic capacity (Delazer et al., 2003). Retrieval of proper names has been linked to activity of the anterior inferior temporal lobe (Gainotti, 2007), which is affected by the spread of tauopathy. Unsurprisingly then, lexical analysis (i.e., analysis of the specific words individuals recall) is an effective marker of AD (Mueller et al., 2020). In particular, failure to delay recall proper names shows strong correlations to AD biomarkers, including PET (Mueller et al., 2022) and CSF (Hale et al., 2024). Additionally, serial position analysis refers to quantifying the serial position curve, a common pattern in tests of human memory where performance tends to be better for stimuli learned at the beginning (primacy) and/or at the end (recency) of a list, as opposed to the middle (Bruno et al., 2013; Hermann et al., 1996; Murdock, 1962). When applying serial position analysis of the LMT to AD biomarkers, loss of primacy recall has been associated with PET- and CSF-determined AD pathology (Bruno et al., 2021, 2023), and forgetting of the entire story was linked to plasma p-tau217 levels (Jauregi-Zinkunegi, Wilson, et al., 2025).

With the present study, we compared a range of LMT metrics against plasma AD biomarkers to determine the most effective at detecting *in vivo* levels of pathology, cross-sectionally. Cross-sectional analyses, compared to longitudinal analyses, provide information about what may be clinically relevant to assess in order to establish current level of pathology. We carried out *specification curve analysis* (SCA; 38) on data from the Wisconsin Registry of Alzheimer's Prevention (WRAP). This approach allows to run various (plausible) models with different permutations of covariates while maintaining predictor and outcome constant (Steegen et al., 2016), which allows us to ascertain how often a certain predictor is significant across permutations, thus giving us an estimate of its robustness. In these tests, predictors were LMT metrics, and outcomes were the plasma levels of p-tau217; amyloid-β 1–40 (Aβ40), amyloid-β 1–42 (Aβ42), and their ratio (Aβ42/40); neurofilament light (NfL), which measures axonal damage (Götze et al., 2024); and glial fibrillary acidic protein (GFAP), which is a marker of reactive astrogliosis (Benedet et al., 2021). To note, this study is analogous to the one we previously conducted in this same cohort (Jauregi-Zinkunegi et al., 2025b) – however, the current examination intends to compare serial position- and forgetting-based metrics to lexical markers directly; employ a larger sample size (416 *vs* 1195); and evaluate potential sex-related differences, given

that women account for more AD cases than men (Nichols et al., 2022).

2. Methods

2.1. Participants

Data were obtained from WRAP (Johnson et al., 2018), an ongoing longitudinal cohort study based at the University of Wisconsin–Madison, USA. Participants are volunteers from the local area. None have a dementia diagnosis or evidence of dementia at enrolment. Participants attend a baseline visit, then are invited for a first follow-up two to four years after baseline and then every two years after that. Participants in our analyses all had to have measures of plasma biomarkers and LMT data. In cases where multiple plasma samples were available for a participant, we selected the most recent sample and the corresponding cognitive assessment. Typically, biomarker extraction and cognitive assessment were concurrent, and when not, we only included cases in which these events were within two years of each other. Participants were only included in our analyses if cognitively unimpaired (stable or declining) or with MCI at the assessment visit of relevance. Cognitive statuses were assessed using a two-tiered consensus conference approach (Langhough et al., 2021), based on core clinical criteria developed by the National Institute on Aging and the Alzheimer's Association (Albert et al., 2011; McKhann et al., 2011). All in all, these criteria left us with 1195 individuals, means and standard deviations of demographic variables are reported in Table 1. Of the total sample, 947 were classified as cognitively unimpaired stable, 181 as cognitively unimpaired declining, and 67 with MCI. Approximately 35% of these participants were included in a previous study from our group with LMT and plasma biomarkers (Jauregi-Zinkunegi et al., 2025b). In addition to a larger sample, the present study also benefits from updated data, a different and more comprehensive analytical approach, a wider comparison of predictors, and sex-differential analyses. All activities for this study were approved by the ethics committees of the authors' universities and conducted in accordance with the Declaration of Helsinki. All participants provided informed consent prior to testing.

Table 1 – Means and standard deviations (except for "cognitive visit number") of the study variables.

	All ($n = 1195$)	Females ($n = 828$)	Males ($n = 367$)
Age at plasma extraction	67.23 (7.68)	67.01 (7.60)	67.74 (7.85)
Elapsed time (years)	.50 (.32)	.51 (.31)	.49 (.32)
APOE risk score	1.17 (.73)	1.20 (.76)	1.10 (.65)
Education (years)	16.16 (2.79)	16.05 (2.75)	16.42 (2.86)
Cognitive visit number (median and range)	6 (1–8)	6 (1–8)	6 (1–8)
Plasma p-tau217	.43 (.32)	.43 (.33)	.43 (.30)
Plasma Aβ42	6.74 (2.20)	6.83 (2.24)	6.52 (2.10)
Plasma Aβ40	103.39 (26.39)	104.28 (26.25)	101.38 (26.63)
Plasma Aβ42/40 ratio	.07 (.02)	.07 (.02)	.07 (.02)
Plasma NfL	19.05 (12.76)	19.30 (13.32)	18.49 (11.41)
Plasma GFAP	134.69 (75.52)	143.42 (78.92)	115.00 (63.03)
LMT immediate recall	13.63 (3.48)	13.85 (3.47)	13.15 (3.47)
LMT delayed recall	12.19 (4.02)	12.43 (4.04)	11.63 (3.91)
LMT primacy ratio	1.27 (.42)	1.25 (.39)	1.31 (.48)
LMT recency ratio	1.07 (.26)	1.08 (.27)	1.07 (.22)
LMT total ratio	1.17 (.33)	1.16 (.30)	1.19 (.40)
LMT delayed proper names	.46 (.25)	.48 (.25)	.42 (.24)

Note. *APOE*: apolipoprotein *E*; p-tau: phosphorylated tau; Aβ: amyloid-beta; GFAP: glial fibrillary acidic protein; NfL: neurofilament light chain; LMT: Logical Memory Task. Plasma biomarker levels in pg/mL.

2.2. Materials

The LMT is a subtest of the Wechsler Memory Scale Revised (WMS-R; 18), and comprised two stories ("A" and "B"), each with 25 items ("idea units"). Each story is read aloud to the participant, and then the participant is asked to recall each story immediately after presentation, and again after a 25–30 min delay. Participants are free to recall the items in any order. Scoring procedures from the WMS-R manual were applied. Two standard metrics were extracted from the LMT (averaging over A and B): the raw score for LMT Immediate recall, derived from the total number of idea units recalled immediately after learning the story; and the raw score for LMT Delayed recall, derived from the total number of idea units recalled after the delay. Process metrics were delayed proper noun (dPN) recall, the primacy ratio (Pr), the recency ratio (Rr), and the total ratio (Tr). dPN recall was calculated by assigning each correct response to delayed LMT a lexical category (e.g., proper names, verbs, numerical expressions) by running a transcript of the LMT stories through CLAN (Computerized Language Analysis Program), a part-of-speech tagger (Macwhinney et al., 2011). Idea units identified as proper names were then summed for stories A and B to create a dPN recall score. A higher dPN score indicated better recall (for previous literature on this method see (Mueller et al., 2022) (Mueller et al., 2025), (Cowman et al., 2025)). To create Pr, Rr and Tr first primacy and recency were defined as the first and final eight idea units of each story. Pr then was given by dividing immediate primacy performance (+1) by delayed primacy performance (+1), averaging across stories. The +1 correction is applied to avoid 0 scores. Rr was given by dividing immediate recency performance (+1) by delayed recency performance (+1), and averaging across stories. And Tr was given by dividing immediate recall (i.e., the whole story) performance (+1) by delayed recall performance (+1), averaging across stories. Higher scores for these ratios indicate more forgetting, hence poorer recall (for previous literature on this method see: 16, 22, 25).

2.3. Blood biomarkers

Plasma samples were analyzed at the Department of Psychiatry and Neurochemistry, University of Gothenburg (Ashton et al., 2024). Plasma Aβ42 and Aβ40, GFAP and NfL were quantified using the Simoa® Human Neurology 4-Plex *E* assay. P-tau217 levels were measured using the Simoa® ALZpath p-tau 217 assay. All plasma levels were analyzed using Quanterix's HD-X™ Analyzer.

2.4. Genotyping

DNA was extracted from whole blood. Samples were aliquoted on 96-well plates for determination of apolipoprotein E (*APOE*) genotypes. An *APOE* risk score was calculated based on the odds ratios of the ε2/ε3/ε4 genotype, as previously reported (Albert et al., 2011).

2.5. Design and analysis

Assumptions of normality and homoscedasticity were checked, along with Q–Q plots. Plasma biomarkers, p-tau217, GFAP, NfL, Aβ42, Aβ40, and the Aβ42/40 ratio, were log10 transformed due to non-normal distributions. Total Immediate recall, Total Delayed recall, Pr, Rr, Tr and Delayed Proper Nouns recall scores were z-standardised to allow their comparison. Bivariate Pearson correlations were conducted between LMT scores, covariates and plasma biomarker levels, for descriptive purposes.

To determine which LMT scores are most effective at detecting plasma biomarker levels, we carried out a SCA, which is a multiverse-based approach that examines all plausible combinations of predictors, outcomes, and covariates to address a given research question (Steegen et al., 2016). This approach offers important advantages, including reducing the overall false-positive rate and strengthening the robustness of findings (Cosme & Lopez, 2023; Flournoy et al., 2020; Klapwijk et al., 2021; Yuan et al., 2022). SCA allows us to explore multiple plausible frequentist analytical choices simultaneously and to quantify how often a predictor remains statistically significant, providing an estimate of how stable and consistent the association is across models.

In the analysis, we kept the outcome–predictor combinations constant and varied model specifications along two dimensions: covariate inclusion and sample subset. For covariates, we tested all 32 possible combinations of the five control variables: age at blood extraction, years of education, *APOE* risk score, cognitive visit number (as practice effects), and the elapsed time between cognitive assessment and blood extraction. For sample subset, models were run separately with 1) the full sample, 2) males only, and 3) females only, based on self-reported sex. Thus, each model specification included one predictor, one outcome, a covariate combination, and one subset, resulting in 6 outcomes × 6 predictors × 32 covariate sets × 3 subsets = 3,456 linear models. Specification curves were used to visualise the full distribution of associations, along with density, forest and percentage of significance plots.

To test the robustness of our findings, we implemented two inferential procedures as recommended by Simonsohn et al. (2020). First, to assess whether the overall pattern of results across the multiverse could have occurred by chance, we conducted a global permutation test. Specifically, a null distribution was generated by repeating the full SCA 1,000 times using permuted data in which the outcome variables were independently shuffled across participants. For each permuted dataset, the total number of significant associations across the multiverse was recorded. The observed number of significant effects in the original (non-shuffled) data was then compared to this null distribution to obtain a global permutation-based *p*-value, which reflects the probability of observing as many or more significant effects under the null hypothesis of no true association. Second, to assess the robustness of individual predictor-outcome combinations, we used two complementary approaches. Using Stouffer's method (Stouffer et al., 1949), *p*-values from all 32 covariate models were aggregated to produce a single *p*-value per predictor-outcome pair. We also computed a permutation-based *p*-value for the median effect size by comparing the observed median to the distribution of medians obtained from the 1,000 permuted SCAs. For all statistical analyses, significance was set as $p < .05$.

Statistical analyses were carried out using *R* software (version 4.4.0). Data processing and visualisation were performed using the *tidyverse* suite of packages. Bivariate correlations were calculated with the *corrr* package, and the specification curve analysis was conducted with the *specr* package (Masur & Scharkow, 2020). All packages are available from the Comprehensive *R* Archive Network (CRAN).

3. Results

In Table 1, means, standard deviations and ranges are described for all the variables included in the current study, for the whole sample, females and males.

3.1. Bivariate correlations

As shown in Table 2, bivariate Pearson correlations between LMT scores and plasma biomarkers were generally small in magnitude. The largest LMT–biomarker correlation was observed between Total ratio and plasma p-tau217 ($r = .157$). Plasma p-tau217 showed the broadest pattern of correlations, with all LMT metrics reaching statistical significance. Several LMT metrics were also significantly correlated with NfL and GFAP, although these correlations were similarly small. Correlations with Aβ42/40, Aβ40, and Aβ42 were present for fewer LMT metrics. Detailed correlation heatmaps for the whole sample, females, and males are provided in Supplementary Fig. S1–S3.

3.2. Specification curve analysis (SCA)

A specification curve analysis was conducted to identify which LMT metrics were most consistently associated with cross-sectional plasma biomarker levels across alternative model specifications. A summary of results, including median effect sizes, confidence intervals and robustness estimates for each significant model, is shown in Table 3. Only the associations that passed the permutation test are shown (permutation $p < .05$), as this ensures a better control against false positives, see Supplementary Table 1 for results with all models. Additional plots, including forest and percentage of significance plots for each outcome and subset can be viewed in Figs. 1 and 2. Fig. S4 shows the specification curve, specifically, standardised regression coefficients for each model specification are shown in panel A, and in panel B, the modelling decisions corresponding to each specification are shown. The global permutation test confirmed that the

Table 2 – Bivariate correlations between plasma biomarkers and Logical Memory Task scores.

Variable	p-tau217	Aβ42	Aβ40	Aβ42/40	NfL	GFAP
Immediate recall	−.150***	.009	−.058*	.070*	−.122***	−.111***
Delayed recall	−.156***	.019	−.037	.060*	−.120***	−.112***
Primacy ratio	.133***	−.054	−.048	−.019	.043	.044
Recency ratio	.075**	−.008	−.009	−.001	.041	.039
Total ratio	.157***	−.053	−.024	−.043	.047	.082**
Del PN	−.155***	.015	−.034	.052	−.107***	−.073*
Age	.259***	.112***	.241***	−.104***	.374***	.467***
Education	−.001	−.035	−.014	−.020	−.012	.047
APOE	.226***	−.180***	−.090**	−.149***	−.061*	.084**
Practice	.161***	.002	.013	−.004	.155***	.233***
Time	−.003	−.002	.027	−.034	−.013	−.034

Note. N = 1195. p-tau: phosphorylated tau; Aβ: amyloid-beta; NfL: neurofilament light chain; GFAP: glial fibrillary acidic protein. *$p < .05$, **$p < .01$, ***$p < .001$.

Table 3 – Summary of median standardised regression coefficients, 95% confidence intervals, and robustness estimates for outcome–predictor combinations that survived permutation testing.

Outcome	Group	Predictor	Median β	Lower CI	Upper CI	% Sig.	Stouffer p	Perm. p
p-tau217	All	Immediate	−.03	−.05	−.02	100	<.001	<.001
		Delayed	−.03	−.05	−.02	100	<.001	<.001
		Tr	.04	.03	.04	100	<.001	<.001
		Rr	.02	.01	.02	100	<.001	.014
		Pr	.03	.02	.04	100	<.001	<.001
		Del PN	−.03	−.05	−.02	100	<.001	<.001
	Female	Immediate	−.04	−.05	−.02	100	<.001	<.001
		Delayed	−.04	−.05	−.02	100	<.001	<.001
		Tr	.04	.03	.05	100	<.001	<.001
		Rr	.02	.01	.02	87.5	<.001	.041
		Pr	.03	.02	.04	100	<.001	<.001
		Del PN	−.03	−.04	−.02	100	<.001	<.001
	Male	Delayed	−.03	−.04	−.02	62.5	<.001	.040
		Tr	.03	.02	.04	100	<.001	.011
		Pr	.02	.02	.03	68.8	<.001	.036
		Del PN	−.04	−.06	−.03	100	<.001	<.001
Aβ40	All	Pr	−.01	−.01	.00	50	<.001	.042
	Female	Tr	−.01	−.01	−.01	100	<.001	.018
		Pr	−.01	−.02	−.01	100	<.001	.001
Aβ42	Female	Tr	−.02	−.02	−.01	100	<.001	.002
		Pr	−.02	−.02	−.01	100	<.001	.005
	Male	Rr	.02	.02	.03	100	<.001	.004
Aβ42/40	All	Immediate	.01	.00	.01	53.1	<.001	.047
NfL	All	Immediate	−.02	−.04	−.01	100	<.001	.012
		Delayed	−.02	−.04	−.01	100	<.001	.009
	Female	Immediate	−.02	−.04	−.02	100	<.001	.015
		Delayed	−.02	−.04	−.01	100	<.001	.028
		Del PN	−.02	−.04	−.01	90.6	<.001	.041
GFAP	All	Immediate	−.02	−.04	−.01	87.5	<.001	.010
		Delayed	−.02	−.04	−.01	87.5	<.001	.009
	Female	Immediate	−.03	−.05	−.02	100	<.001	<.001
		Delayed	−.03	−.05	−.02	100	<.001	<.001
		Tr	.02	.02	.03	100	<.001	.005
		Del PN	−.02	−.04	−.01	100	<.001	.008

number of statistically significant associations observed in the data was greater than expected by chance (permutation-based $p < .001$; see also the histogram in Supplementary Fig. S5). Density plots for each outcome by whole sample, female and males are provided in Supplementary Fig. S6.

Plasma p-tau217. Across the multiverse, the most consistent and stable relationships were found for plasma p-tau217 biomarker levels. All six LMT scores were statistically robust, and as expected, Immediate, Delayed recall, and Delayed PN showed negative associations, while the ratios, i.e., Pr, Rr, and Tr, showed positive associations. The six predictors showed 100% specification significance and remained statistically significant under both permutation and Stouffer p-value thresholds in the full sample.

Subset analyses revealed more stable associations in females, where all six LMT predictors remained statistically

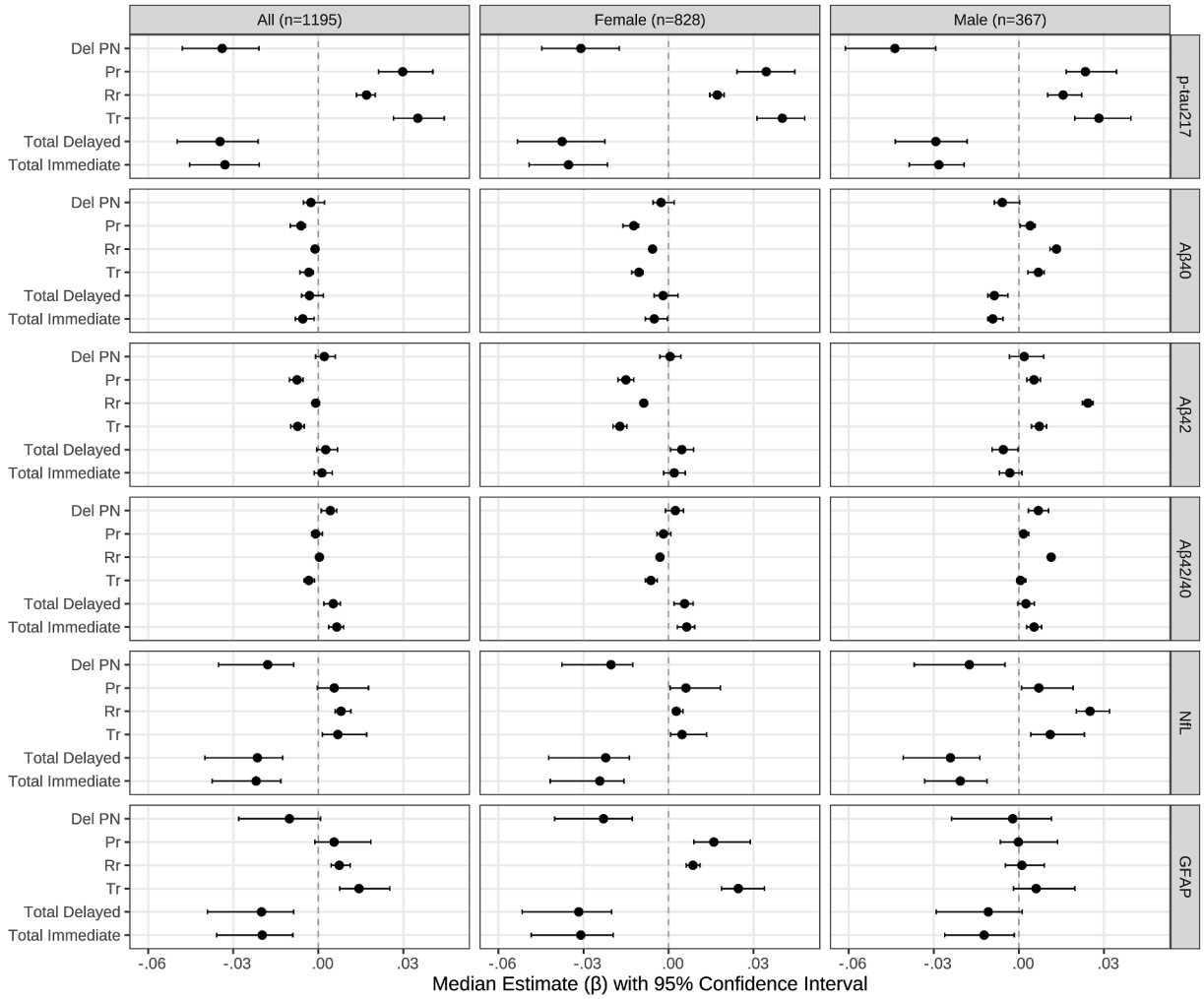

Fig. 1 – Forest plots for each outcome by whole sample, female and males. Note. p-tau: phosphorylated tau; Aβ: amyloid-beta; GFAP: glial fibrillary acidic protein; NfL: neurofilament light chain; LMT: Logical Memory Task; Tr: Total ratio; Rr: Recency ratio; Pr: Primacy ratio; Del PN: Delayed Proper Nouns.

robust and showed consistency across specifications. In males, Delayed recall, Pr, Tr, and Delayed PN were significant, yet specification consistency was low for all except Tr and Delayed PN.

Plasma Aβ42/40 ratio. As Aβ40 is meant to provide a control value for Aβ42, rather than a signal of AD pathology *per se*, we focus here on the ratio results, but leave both Aβ40 and Aβ42 values in Table and Figures for full illustrative purposes. In the full sample, only Immediate recall showed a weak but statistically robust positive association, yet consistency across model specifications was of 53%. Stratified analyses revealed no consistent or robust predictors in female or male subgroups.

Plasma NfL. Consistent and statistically robust associations were observed between LMT scores and plasma NfL levels. In the full sample, Immediate and Delayed recall showed robust negative associations, and although Delayed PN was also negatively associated, it did not survive the permutation threshold. These associations were strongest and most robust in females, where all three (Immediate, Delayed, and Delayed PN) were consistent across specifications and remained statistically robust. In males, the three showed significant associations in 50% of model specifications, but none met robustness criteria. Only Rr showed a high proportion of significant associations, but it did not survive the permutation test.

Plasma GFAP. Several LMT scores were associated with GFAP. In the full sample, Immediate and Delayed recall showed a high proportion of significant and robust negative associations. In females, all LMT predictors except Rr and Pr were significantly and robustly associated with GFAP, showing significant associations in 100% of model specifications. This pattern was not found in males, where the specification significance was low and no associations survived permutation-based significance.

Overall, the SCA indicated that associations with p-tau217 were the most stable across LMT metrics and model specifications. Associations with NfL and GFAP were also observed

Fig. 2 – Plots of percentages of significance for each outcome by whole sample, female and males. Note. p-tau: phosphorylated tau; Aβ: amyloid-beta; GFAP: glial fibrillary acidic protein; NfL: neurofilament light chain; LMT: Logical Memory Task; Tr: Total ratio; Rr: Recency ratio; Pr: Primacy ratio; Del PN: Delayed Proper Nouns.

for some LMT metrics but varied more by metric and subgroup, while associations with Aβ measures were less consistent.

4. Discussion

The purpose of this investigation was to compare a range of story recall metrics against plasma AD biomarkers to determine which are the most sensitive to *in vivo* levels of pathology, cross-sectionally. To this end, we carried out multiverse analysis/SCA, which allowed us to look at a variety of alternative analytical choices simultaneously. Predictors were LMT-derived standard and process-based metrics, while outcomes were p-tau217, the Aβ42/40 ratio, NfL and GFAP. Within the full sample, we observed that associations were the most stable and predictable between LMT scores and p-tau217, as all metrics predicted biomarker levels significantly 100% of the time. In contrast, associations between LMT scores and other plasma biomarkers were generally less reliable.

Plasma p-tau217 has been at the forefront of research on blood-based biomarkers of early detection of AD, due to its widely reported sensitivity and robustness (Arranz et al., 2025), as previously discussed, therefore it is not surprising that its levels should be strongly correlated with cross-sectional memory performance, particularly when examining the full sample. In contrast, research has shown plasma Aβ markers, including the Aβ42/40 ratio, to be generally less sensitive in comparison (Jack et al., 2024) – possibly owing to their reduced stability across time, lab-based variables and patient-related factors (Piura et al., 2025). Our findings are consistent with these claims: amyloid-β levels did not significantly associate with any story recall metric as often as p-tau217 levels did, at least in the full group. Finally, NfL and GFAP are effective biomarkers of neurodegeneration across a range of pathologies (Pilotto et al., 2024) rather than being AD-specific. Our results showed that NfL and GFAP are consistently predicted in the full cohort by traditional LMT scores, immediate and delayed recall, rather than by process scores—this finding suggests that traditional metrics may be

preferable for detecting overall and unspecific damage to the cognitive infrastructure.

Women account for more AD cases than men (Nichols et al., 2022), as such we wanted to evaluate differences in the associations between LMT metrics and plasma biomarkers across sexes. One caveat is that the female group was larger ($n = 828$) than the male group ($n = 367$), so differences may also be due to relative power, although the male group was still sizeable. Examining associations between p-tau217 and LMT metrics in females, we found that the associations were significant 100% of the time except for Rr (reaching 87.5%). However, in males, only delayed PN recall (Basche et al., 2024) and Tr (consistent with our previous report, 27) were predictive 100% of the time. Similar sex-differential associations between memory performance and plasma p-tau217 have been reported previously (Saloner et al., 2024). Notably, in our data, the levels of p-tau217 across groups were analogous (Table 1), presumably discounting differences in the levels of pathology. Relative power may play a role, as noted, suggesting that delayed PN recall and Tr, possibly having stronger cross-sectional associations with p-tau217, are more resistant to smaller sample sizes. In this respect, Fig. S5 provides a visual illustration of how these two LMT metrics' coefficients reach the out the farthest to the left and the right, respectively, denoting the highest effect sizes (in males). Hence, these results suggest that using delayed PN recall and Tr as a proxy of p-tau 217, regardless of gender, could yield the most sensitive results. Alternatively, AD pathology may have affected memory less in the male sub-group compared to the female sub-group, perhaps due the slightly higher level of education or by the lower *APOE* risk reported by the former (Table 1) – although, of course, education and *APOE* risk were included as covariates, and thus controlled for, in some of the model permutations.

In general, across the remaining outcomes, associations were more robust and consistent in females than in males, as illustrated in Table 3 and Fig. 2. This discrepancy is especially visible with GFAP, where levels are higher in females than males in our sample (Table 1), and there are reliable associations between GFAP and delayed PN recall, Tr, and delayed and immediate LMT recall in females (100% of the time across permutations), but only weak associations in males. A recent report also echoes these findings (Plaska et al.) and suggests that increased astrocytic activation may underline at least in part the increased risk of neurodegeneration in women. Considering these and other findings, we conducted *post-hoc* analyses to determine whether the association between p-tau217 and LMT metrics was moderated by plasma GFAP levels and/or sex. Our results showed that plasma p-tau217 and GFAP interacted in predicting delayed LMT ($p = .046$), Pr ($p = .005$), Tr ($p = .010$), and delayed PN recall ($p = .036$), but none of the three-way interactions with sex were significant ($p_s > .078$). These *post hoc* analyses emphasize the important role of astrocyte activity and inflammation in conjunction with tauopathy for predicting memory decline, although they are partial and do not inform on directionality. These results do not directly implicate sex, suggesting that elevated GFAP may be deleterious regardless of sex. However, as women tend to have on average higher plasma GFAP levels (potentially due to multiple factors), this sex-related discrepancy may help explain why plasma p-tau217 levels correlate more strongly with cognitive performance in this group compared to males. All in all, we join many researchers in the AD field advocating for a deeper exploration of these differences (Buckley et al., 2023).

A strength of this study is that we employed a Multiverse/SCA approach which allows to explore multiple plausible analytical choices simultaneously—in other words, we were able to compare multiple LMT metrics, against multiple plasma biomarkers, and determine quantitatively how frequently an association was significant across permutations of covariates. In this way, we can determine conclusively which metrics consistently predict biomarker outcomes, under various conditions. However, a key limitation is that the total sample, only less than 8% of the participants identified as non-white, thus emphasising a lack of ethnic variety, which is unfortunately common in AD research (Gilmore-Bykovskyi et al., 2019). The small number of non-white participants also preclude the possibility of examining differences across racial or ethnic groups. Future research should delve further into the sex-related differences in AD pathology, biomarkers levels, and cognitive performance outputs in a more diverse sample.

To conclude, in this multiverse analysis of 1195 WRAP participants, LMT scores and plasma biomarkers, we observed the following: p-tau217 levels were most consistently associated with memory performance, compared to other biomarkers; there were clear sex differences in the strength of these associations; and delayed recall of proper nouns (dPN) and overall forgetting (Tr) were the most reliable predictors of plasma p-tau217 levels in this cohort. The final point further suggests that, when using LMT (and including when evaluating ADNI data), regardless of sex, but in particular with males, calculating and evaluating dPN and Tr may provide sensitive cognitive data to evaluate a person's likelihood of underlying AD pathology. The development of metrics able to signal subtle impairment, beyond the traditional diagnostic cutlines, is key to revealing phenotypic heterogeneity in otherwise cognitively-unimpaired subjects (Kiselica & Working Group on Objective Subtle Cognitive Decline in Alzheimer's Disease, 2026), and will aid accounting for subtle differences that emerge during the preclinical phases of AD.

Consent statements

This study was based on secondary data analysis. The original data were collected after obtaining consent from study participants.

Funding sources

This secondary analysis of WRAP data was funded by a NIH-NIA (R01AG07940-01) grant to KDM, in which DB is co-investigators, and AJZ is postdoc. Primary data collection for WRAP data was supported by NIH-NIA grants 5R01AG027161-13 and 5R01AG021155-18 on which SCJ is principal investigator.

HZ is a Wallenberg Scholar and a Distinguished Professor at the Swedish Research Council supported by grants from the

Swedish Research Council (#2023-00356, #2022-01018 and #2019–02397), the European Union's Horizon Europe research and innovation programme under grant agreement No 101053962, Swedish State Support for Clinical Research (#ALFGBG-71320), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809–2016862), the AD Strategic Fund and the Alzheimer's Association (#ADSF-21-831376-C, #ADSF-21-831381-C, #ADSF-21-831377-C, and #ADSF-24-1284328-C), the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States (NEuroBioStand, #22HLT07), the Bluefield Project, Cure Alzheimer's Fund, the Olav Thon Foundation, the Erling-Persson Family Foundation, Familjen Rönströms Stiftelse, Familjen Beiglers Stiftelse, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2022-0270), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), the European Union Joint Programme–Neurodegenerative Disease Research (JPND2021-00694), the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre, the UK Dementia Research Institute at UCL (UKDRI-1003), and an anonymous donor.

Declaration of competing interest

HZ has served at scientific advisory boards and/or as a consultant for Abbvie, Acumen, Alector, Alzinova, ALZpath, Amylyx, Annexon, Apellis, Artery Therapeutics, AZTherapies, Cognito Therapeutics, CogRx, Denali, Eisai, Enigma, LabCorp, Merck Sharp & Dohme, Merry Life, Nervgen, Novo Nordisk, Optoceutics, Passage Bio, Pinteon Therapeutics, Prothena, Quanterix, Red Abbey Labs, reMYND, Roche, Samumed, ScandiBio Therapeutics AB, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures sponsored by Alzecure, BioArctic, Biogen, Cellectricon, Fujirebio, LabCorp, Lilly, Novo Nordisk, Oy Medix Biochemica AB, Roche, and WebMD, is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program, and is a shareholder of CERimmune Therapeutics (outside submitted work).

No other conflicts of interest are reported.

Acknowledgements

We would like to acknowledge WRAP study participants and staff for their invaluable contributions to this research. We also acknowledge the entire staff of the Cognitive-Communication in Aging and Neurogenic Disorders Laboratory (CCANDL) at University of Wisconsin–Madison for their efforts in coding and managing the data for this study.

Scientific transparency statement

DATA: Some raw and processed data supporting this research are publicly available, while some are subject to restrictions: https://wrap.wisc.edu/data-requests-2/

CODE: All analysis code supporting this research is publicly available: https://osf.io/k8amh.

MATERIALS: This research did not make use of any materials to generate or acquire data.

DESIGN: This article reports, for all studies, how the author(s) determined all sample sizes, all data exclusions, all data inclusion and exclusion criteria, and whether inclusion and exclusion criteria were established prior to data analysis.

PRE-REGISTRATION: No part of the study procedures was pre-registered in a time-stamped, institutional registry prior to the research being conducted. No part of the analysis plans was pre-registered in a time-stamped, institutional registry prior to the research being conducted.

For full details, see the *Scientific Transparency Report* in the supplementary data to the online version of this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2026.06.005.

REFERENCES

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia, 7*(3), 270–279.

Ansu-Mensah, M., Bawontuo, V., Kuupiel, D., & Ginindza, T. G. (2024). Sustainable solutions to barriers of point-of-care diagnostic testing services in health facilities without laboratories in the bono region, Ghana: A qualitative study. *BMC Primary Care, 25*(1), Article 179.

Arranz, J., Ferrer, R., Zhu, N., Rubio-Guerra, S., Rodríguez-Baz, Í., Arriola-Infante, J. E., Maure-Blesa, L., Garcia-Castro, J., Selma-González, J., Carmona-Iragui, M., & Barroeta, I. (2025). Prospective evaluation of plasma pTau217 stability for the detection of Alzheimer's disease in a tertiary memory clinic. *Alzheimer's Research and Therapy, 17*(1), Article 150.

Ashton, N. J., Brum, W. S., Di Molfetta, G., Benedet, A. L., Arslan, B., Jonaitis, E., Langhough, R. E., Cody, K., Wilson, R., Carlsson, C. M., Vanmechelen, E., Montoliu-Gaya, L., Lantero-Rodriguez, J., Rahmouni, N., Tissot, C., Stevenson, J., Servaes, S., Therriault, J., Pascoal, T., … Zetterberg, H. (2024). Diagnostic accuracy of a plasma phosphorylated tau 217 immunoassay for Alzheimer disease pathology. *JAMA Neurology, 81*(3), 255–263.

Baek, M. J., Kim, H. J., Ryu, H. J., Lee, S. H., Han, S. H., Na, H. R., Chang, Y., Chey, J. Y., & Kim, S. (2011). The usefulness of the story recall test in patients with mild cognitive impairment and Alzheimer's disease. *Aging, Neuropsychology, and Cognition, 18*(2), 214–229.

Balogun, W. G., Zetterberg, H., Blennow, K., & Karikari, T. K. (2023). Plasma biomarkers for neurodegenerative disorders: Ready for prime time? *Current Opinion in Psychiatry, 36*(2), 112–118.

Basche, K. E., Langhough, R. E., Jonaitis, E. M., Du, L., Bruno, D., Wilson, R. E., Hermann, B. P., Johnson, S. C., & Mueller, K. D.

(2024). Longitudinal change in proper names from logical memory tasks are sensitive to plasma pTau217 levels. *Alzheimer's and Dementia, 20*, Article e087520.
Bellass, S., Canvin, K., & Sheard, L. (2024). 'Trying to battle a very slow version of the system that exists outside': Experiences of waiting for healthcare in English prisons. *Health, 28*(5), 736–753.
Benedet, A. L., Milà-Alomà, M., Vrillon, A., Ashton, N. J., Pascoal, T. A., Lussier, F., Karikari, T. K., Hourregue, C., Cognat, E., Dumurgier, J., & Stevenson, J. (2021). Differences between plasma and cerebrospinal fluid glial fibrillary acidic protein levels across the Alzheimer disease continuum. *JAMA Neurology, 78*(12), 1471–1483.
Bruno, D., Jauregi Zinkunegi, A., Kollmorgen, G., Carboni, M., Wild, N., Carlsson, C., Bendlin, B., Okonkwo, O., Chin, N., Hermann, B. P., Asthana, S., Blennow, K., Langhough, R., Johnson, S. C., Pomara, N., Zetterberg, H., & Mueller, K. D. (2023). A comparison of diagnostic performance of word-list and story recall tests for biomarker-determined Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 45*(8), 763–769.
Bruno, D., Jauregi-Zinkunegi, A., Betthauser, T., Carlsson, C., Bendlin, B. B., Okonkwo, O., Chin, N. A., Asthana, S., Langhough, R. E., Johnson, S. C., & Mueller, K. D. (2024). Story recall performance and AT classification *via* positron emission tomography: A comparison of logical memory and craft story 21. *Journal of the Neurological Sciences, 464*, Article 123148.
Bruno, D., Mueller, K. D., Betthauser, T., Chin, N., Engelman, C. D., Christian, B., Koscik, R. L., & Johnson, S. C. (2021). Serial position effects in the logical memory test: Loss of primacy predicts amyloid positivity. *Journal of Neuropsychology, 15*(3), 448–461.
Bruno, D., Reiss, P. T., Petkova, E., Sidtis, J. J., & Pomara, N. (2013). Decreased recall of primacy words predicts cognitive decline. *Archives of Clinical Neuropsychology, 28*(2), 95–103.
Buckley, R. F., Gong, J., & Woodward, M. (2023). A call to action to address sex differences in Alzheimer disease clinical trials. *JAMA Neurology, 80*(8), 769–770.
Cosme, D., & Lopez, R. B. (2023). Neural indicators of food cue reactivity, regulation, and valuation and their associations with body composition and daily eating behavior. *Social Cognitive and Affective Neuroscience, 18*(1), Article nsaa155.
Cowman, D., Langhough, R., Olson, H., Basche, K., Sanson-Miles, L., Bruno, D., Hermann, B., Christian, B. T., Betthauser, T. J., Johnson, S. C., & Mueller, K. D. (2025). Sex differences in story recall decline in preclinical Alzheimer's disease. *Brain Communications, 7*(3), Article fcaf169.
Delazer, M., Semenza, C., Reiner, M., Hofer, R., & Benke, T. (2003). Anomia for people names in DAT—Evidence for semantic and post-semantic impairments. *Neuropsychologia, 41*(12), 1593–1598.
Della Monica, C., Revell, V., Atzori, G., Laban, R., Skene, S. S., Heslegrave, A., Hassanin, H., Nilforooshan, R., Zetterberg, H., & Dijk, D. J. (2024). P-tau217 and other blood biomarkers of dementia: Variation with time of day. *Translational Psychiatry, 14*(1), Article 373.
Duff, K., & investigators, A. D. N. I. (2021). Amnestic MCI in ADNI: Maybe not enough memory impairment? *Neurology, 97*(12), 595–596.
Duff, K., Leber, W. R., Patton, D. E., Schoenberg, M. R., Mold, J. W., Scott, J. G., & Adams, R. L. (2007). Modified scoring criteria for the RBANS figures. *Applied Neuropsychology, 14*(2), 73–83.
Flournoy, J. C., Vijayakumar, N., Cheng, T. W., Cosme, D., Flannery, J. E., & Pfeifer, J. H. (2020). Improving practices and inferences in developmental cognitive neuroscience. *Developmental Cognitive Neuroscience, 45*, Article 100807.
Götze, K., Vrillon, A., Dumurgier, J., Indart, S., Sanchez-Ortiz, M., Slimi, H., Raynaud-Simon, A., Cognat, E., Martinet, M., Zetterberg, H., & Blennow, K. (2024). Plasma neurofilament light chain as prognostic marker of cognitive decline in neurodegenerative diseases, a clinical setting study. *Alzheimer's Research and Therapy, 16*(1), Article 231.
Gainotti, G. (2007). Different patterns of famous people recognition disorders in patients with right and left anterior temporal lesions: A systematic review. *Neuropsychologia, 45*(8), 1591–1607.
Gilmore-Bykovskyi, A. L., Jin, Y., Gleason, C., Flowers-Benton, S., Block, L. M., Dilworth-Anderson, P., Barnes, L. L., Shah, M. N., & Zuelsdorff, M. (2019). Recruitment and retention of underrepresented populations in Alzheimer's disease research: A systematic review. *Alzheimer's and Dementia: Translational Research & Clinical Interventions, 5*, 751–770.
Hale, M. R., Langhough, R., Du, L., Hermann, B. P., Van Hulle, C. A., Carboni, M., Kollmorgen, G., Basche, K. E., Bruno, D., Sanson-Miles, L., Jonaitis, E. M., Chin, N. A., Okonkwo, O. C., Bendlin, B. B., Carlsson, C. M., Zetterberg, H., Blennow, K., Betthauser, T. J., Johnson, S. C., & Mueller, K. D. (2024). Associations between recall of proper names in story recall and CSF amyloid and tau in adults without cognitive impairment. *Neurobiology of Aging, 133*, 87–98.
Hermann, B. P., Seidenberg, M., Wyler, A., Davies, K., Christeson, J., Moran, M., & Stroup, E. (1996). The effects of human hippocampal resection on the serial position curve. *Cortex, 32*(2), 323–334.
Jack, J.,C. R., Wiste, H. J., Algeciras-Schimnich, A., Weigand, S. D., Figdore, D. J., Lowe, V. J., Vemuri, P., Graff-Radford, J., Ramanan, V. K., Knopman, D. S., & Mielke, M. M. (2024). Comparison of plasma biomarkers and amyloid PET for predicting memory decline in cognitively unimpaired individuals. *Alzheimer's and Dementia, 20*(3), 2143–2154.
Jauregi-Zinkunegi, A., Betthauser, T., Carlsson, C. M., Bendlin, B. B., Okonkwo, O., Chin, N. A., Asthana, S., Langhough, R. E., Johnson, S. C., Mueller, K. D., & Bruno, D. (2025a). Delayed primacy recall in AVLT is associated with medial temporal tau PET burden in cognitively unimpaired adults. *Cortex, 184*, 47–57.
Jauregi-Zinkunegi, A., Wilson, R. E., Langhough, E., Ashton, N. J., Blennow, K., Johnson, S. C., Zetterberg, H., Bruno, D., & Mueller, K. D. (2025b). Associations between the logical memory test story recall metrics and plasma biomarkers for Alzheimer's disease in individuals free of dementia. *Clinical Neuropsychologist*, 1–20.
Johnson, S. C., Koscik, R. L., Jonaitis, E. M., Clark, L. R., Mueller, K. D., Berman, S. E., Bendlin, B. B., Engelman, C. D., Okonkwo, O. C., Hogan, K. J., & Asthana, S. (2018). The Wisconsin registry for Alzheimer's prevention: A review of findings and current directions. *Alzheimer's and Dementia: Diagnosis, Assessment & Disease Monitoring, 10*, 130–142.
Karikari, T. K. (2022). Blood tests for Alzheimer's disease: Increasing efforts to expand and diversify research participation is critical for widespread validation and acceptance. *Journal of Alzheimer's Disease, 90*(3), 967–974.
Karikari, T. K., Ashton, N. J., Brinkmalm, G., Brum, W. S., Benedet, A. L., Montoliu-Gaya, L., Lantero-Rodriguez, J., Pascoal, T. A., Suárez-Calvet, M., Rosa-Neto, P., & Blennow, K. (2022). Blood phospho-tau in Alzheimer disease: Analysis, interpretation, and clinical utility. *Nature Reviews Neurology, 18*(7), 400–418.
Khachaturian, Z., Carrillo, M. C., & Khachaturian, A. S. (2025). Perspective: Emerging challenges for a future ADNI. *Alzheimer's and Dementia, 21*(2), Article e14534.

Kiselica, A. M., & Working Group on Objective Subtle Cognitive Decline in Alzheimer's Disease. (2026). *Why "normal" cognition is hindering preclinical Alzheimer's trials*. Dementia Researcher. https://www.dementiaresearcher.nihr.ac.uk/blog-why-normal-cognition-is-hindering-preclinical-alzheimers-trials/.

Klapwijk, E. T., van den Bos, W., Tamnes, C. K., Raschle, N. M., & Mills, K. L. (2021). Opportunities for increased reproducibility and replicability of developmental neuroimaging. *Developmental Cognitive Neuroscience, 47*, Article 100902.

Langhough, K., Langhough Koscik, R., Hermann, B. P., Allison, S., Clark, L. R., Jonaitis, E. M., Mueller, K. D., Betthauser, T. J., Christian, B. T., Du, L., Okonkwo, O., Birdsill, A., Chin, N., Gleason, C., & Johnson, S. C. (2021). Validity evidence for the research category, "cognitively unimpaired - declining," as a risk marker for mild cognitive impairment and Alzheimer's disease. *Frontiers in Aging Neuroscience, 13*, Article 688478.

Libon, D. J., Swenson, R., Lamar, M., Price, C. C., Baliga, G., Pascual-Leone, A., Au, R., Cosentino, S., & Andersen, S. L. (2022). The Boston process approach and digital neuropsychological assessment: Past research and future directions. *Journal of Alzheimer's Disease, 87*(4), 1419–1432.

Macwhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*, 1286–1307.

Masur, P. K., & Scharkow, M. (2020). *Specr: Conducting and visualizing specification curve analyses*.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). 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. *Alzheimer's and Dementia, 7*(3), 263–269.

Mueller, K. D., Du, L., Bruno, D., Betthauser, T., Christian, B., Johnson, S., Hermann, B., & Koscik, R. L. (2022). Item-level story recall predictors of amyloid-beta in late middle-aged adults at increased risk for Alzheimer's disease. *Frontiers in Psychology, 13*, Article 908651.

Mueller, K. D., Koscik, R. L., Du, L., Bruno, D., Jonaitis, E. M., Koscik, A. Z., Christian, B. T., Betthauser, T. J., Chin, N. A., Hermann, B. P., & Johnson, S. C. (2020). Proper names from story recall are associated with beta-amyloid in cognitively unimpaired adults at risk for Alzheimer's disease. *Cortex, 131*, 137–150.

Mueller, K. D., Solden, A., Langhough, R., Bruno, D., Jauregi-Zinkunegi, A., Basche, K., Hale, M., He, D., Moghekar, A., Hermann, B., Albert, M., & Pettigrew, C. (2025). Proper name recall as an early indicator of preclinical Alzheimer's disease pathology. *medRxiv, 39*, 269–276.

Murdock, J.,B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), Article 482.

Nester, C. O., Qin, J., Wang, C., Katz, M. J., Lipton, R. B., & Rabin, L. (2023). Story memory in Alzheimer's disease research: Concordance between logical memory and craft story 21 in a demographically diverse aging cohort. *Alzheimer's and Dementia, 19*, Article e064178.

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., & Al Hamad, H. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the global burden of disease study 2019. *Lancet Public Health, 7*(2), e105–e125.

Palmqvist, S., Warmenhoven, N., Anastasi, F., Pilotto, A., Janelidze, S., Tideman, P., Stomrud, E., Mattsson-Carlgren, N., Smith, R., Ossenkoppele, R., Tan, K., Dittrich, A., Skoog, I., Zetterberg, H., Quaresima, V., Tolassi, C., Höglund, K., Brugnoni, D., Puig-Pijoan, A., … Hansson, O. (2025). Plasma phospho-tau217 for Alzheimer's disease diagnosis in primary and secondary care using a fully automated platform. *Nature Medicine, 31*(6), 2036–2043.

Pandey, N., Yang, Z., Cieza, B., Reyes-Dumeyer, D., Kang, M. S., Montesinos, R., Soto-Añari, M., Custodio, N., Honig, L. S., & Tosto, G. (2025). Plasma phospho-tau217 as a predictive biomarker for Alzheimer's disease in a large South American cohort. *Alzheimer's Research and Therapy, 17*(1), Article 1.

Pilotto, A., Ashton, N. J., Lupini, A., Battaglio, B., Zatti, C., Trasciatti, C., Gipponi, S., Cottini, E., Grossi, I., Salvi, A., & de Petro, G. (2024). Plasma NfL, GFAP, amyloid, and p-tau species as prognostic biomarkers in Parkinson's disease. *Journal of Neurology, 271*(12), 7537–7546.

Piura, Y. D., Figdore, D. J., Lachner, C., Bornhorst, J., Algeciras-Schimnich, A., Graff-Radford, N. R., & Day, G. S. (2025). Diagnostic performance of plasma p-tau217 and Aβ42/40 biomarkers in the outpatient memory clinic. *Alzheimer's and Dementia, 21*(6), Article e70316.

C.R. Plaska, T. Jacobs, D. Bruno, S. Lee, R. Osorio, B.P. Imbimbo, N. Ashton, H. Zetterberg, K. Blennow, N. Pomara (In Preparation). Sex-differences in plasma GFAP and its associations with plasma AD biomarkers and cognition in cognitively unimpaired community-dwelling older adults.

Rozenblatt, S. (2018). Boston process approach. In *Encyclopedia of clinical neuropsychology* (pp. 615–616). Cham: Springer.

Saloner, R., VandeVrede, L., Asken, B. M., Paolillo, E. W., Gontrum, E. Q., Wolf, A., Lario-Lago, A., Milà-Alomà, M., Triana-Baltzer, G., Kolb, H. C., & Dubal, D. B. (2024). Plasma phosphorylated tau-217 exhibits sex-specific prognostication of cognitive decline and brain atrophy in cognitively unimpaired adults. *Alzheimer's and Dementia, 20*(1), 376–387.

Sarfati, M., Senequier, A., Bihouix, P., Marrauld, L., & Manet, R. (2024). Preventing a global health care systems collapse through low-tech medicine. *Journal of Global Health, 14*, Article 03035.

Semenza, C., Mondini, S., Borgo, F., Pasini, M., & Sgaramella, M. T. (2003). Proper names in patients with early Alzheimer's disease. *Neurocase, 9*(1), 63–69.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*(11), 1208–1214.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams Jr, R. M. (1949). The American soldier: Adjustment during army life. *studies in social psychology in world war ii), 1*.

Terao, C. M., Alexander, M. W., Chalmers, R. P., Boshmaf, S. Z., Paterson, J., Black, S. E., Papp, K. V., Sperling, R. A., & Rabin, J. S. (2025). Identifying cognitive test scores associated with early tau burden in Alzheimer's disease. *Alzheimer's and Dementia: Diagnosis, Assessment & Disease Monitoring, 17*(1), Article e70052.

VandeVrede, L., Cho, H., Sanderson-Cimino, M., Wekselman, F., Cobigo, Y., Gorno-Tempini, M. L., Heuer, H. W., Kramer, J. H., Lago, A. L., Leichter, D., & Ljubenkov, P. (2025). Detection of Alzheimer neuropathology in Alzheimer and non-Alzheimer clinical syndromes with blood-based biomarkers. *JAMA Neurology, 82*(4), 344–354.

Wechsler, D. (1987). *WMS-R: Wechsler memory scale-revised*. Psychological Corporation.

Xiao, Z., Zhou, X., Zhao, Q., Cao, Y., & Ding, D. (2024). Significance of plasma p-tau217 in predicting long-term dementia risk in older community residents: Insights from machine learning approaches. *Alzheimer's and Dementia, 20*(10), 7037–7047.

Yakoub, Y., Gonzalez-Ortiz, F., Ashton, N. J., Déry, C., Strikwerda-Brown, C., St-Onge, F., Ourry, V., Schöll, M., Geddes, M. R., Ducharme, S., Montembeault, M., Rosa-Neto, P., Soucy, J. P.,

Breitner, J. C. S., Zetterberg, H., Blennow, K., Poirier, J., Villeneuve, S., & PREVENT–AD Research Group. (2025). Plasma p-tau217 identifies cognitively normal older adults who will develop cognitive impairment in a 10-year window. *Alzheimer's and Dementia, 21*(2), Article e14537.

Yuan, Q., Li, H., Du, B., Dang, Q., Chang, Q., Zhang, Z., Zhang, M., Ding, G., Lu, C., & Guo, T. (2022). The cerebellum and cognition: Further evidence for its role in language control. *Cerebral Cortex, 33*(1), 35–49.

Zhong, X., Wang, Q., Yang, M., Lin, G., Yao, K., Wu, Z., Xu, D., Zhou, H., Chen, B., Shi, H., & Zhang, M. (2025). Plasma p-tau217 and p-tau217/Aβ1-42 are effective biomarkers for identifying CSF-and PET imaging-diagnosed Alzheimer's disease: Insights for research and clinical practice. *Alzheimer's and Dementia, 21*(2), Article e14536.