Digital cognitive test performance and relationship to plasma biomarkers in Alzheimer's disease

Toniolo S[1,2*#], Zhao S[3*#], Scholcz A[3], Amein B[1], Ganse-Dumrath A[1], Heslegrave A J [6,7], Thompson S[2], Manohar S[1,2,3], Zetterberg H[4,5,6,7,8,9], Husain M[1,2,3#]

* Joint first authors, # Corresponding authors

1 Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom.

2 Cognitive Disorders Clinic, JR Hospital, Oxford, United Kingdom.

3 Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom.

4 Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden.

5 Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden.

6 UK Dementia Research Institute, UCL, London, United Kingdom.

7 Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, United Kingdom.

8 Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China.

9 Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA

ORCiD

Sofia Toniolo: https://orcid.org/0000-0002-1833-4995

Sijia Zhao: https://orcid.org/0000-0002-6246-0702

Ganse-Dumrath Akke: https://orcid.org/0000-0002-4828-8117

Sanjay Manohar: http://orcid.org/0000-0003-0735-4349

Henrik Zetterberg: https://orcid.org/0000-0003-3930-4354

Masud Husain: https://orcid.org/0000-0002-6850-9255

Keywords

Cognition, phospho tau, memory, dementia, online testing

Abstract

A major limitation in current Alzheimer's disease (AD) research is the lack of the ability to measure cognitive performance at scale—robustly, remotely, and frequently. For the purposes of screening, recruitment, stratification, and longitudinal follow-up in clinical trials, there are no established online digital platforms validated against plasma biomarkers of AD. Here we report findings using a novel web-based platform that assessed different cognitive functions in AD patients (N=46) and elderly controls (N=53) who were also assessed for plasma biomarkers of AD (amyloid β42/40 ratio, pTau181, GFAP, NfL). Their performance was compared to a second, larger group of elderly controls (N=256). AD patients were significantly impaired across all digital cognitive tests, with performance correlating with plasma biomarker levels, particularly pTau181. These findings show how online testing can now be deployed in AD patients to measure cognitive function effectively and related to blood biomarkers of the disease.

Introduction

The advent of new disease-modifying treatments for Alzheimer's disease (AD) has highlighted the need for sensitive cognitive tests to stratify those who might benefit from early treatment, as well as to track the effects of interventions [1,2]. Traditional face-to-face neuropsychological assessments are able to detect changes only several years after pathological accumulation of amyloid and tau, a factor that is considered crucial for clinical trial failure [3]. Digital metrics obtained using computerised tests can detect subtle signs of impairment that cannot be captured by standard tests of cognition and might be particularly valuable tools in the early phases of the disease, when cognitive impairment is at subthreshold levels on commonly used clinical scales [4]. In principle, such measures could also potentially track disease progression more sensitively [5,6].

Deployment of digital cognitive testing platforms has the potential to make a deep impact on this field, where a crucial limitation has been the lack of the ability to measure cognitive performance at scale [7–10]. Screening for AD, recruitment and stratification into clinical studies, as well as longitudinal follow-up in trials, could all be transformed if cognitive testing were to be conducted robustly, remotely, and frequently [11]. However, any digital cognitive platform first needs to be validated in patients who have evidence of AD pathology, ideally using scalable and affordable biomarkers that might be used in combination to allow for widespread screening or recruitment.

Until quite recently, cerebrospinal fluid (CSF) biomarkers, such as amyloid β (Aβ42 or the Aβ42/40 ratio) and phosphorylated tau 181 (pTau181), have been the major fluid indices used as proxies of AD pathology in the brain. However, work on plasma biomarkers of AD has advanced rapidly. For example, *plasma* Aβ42/40 ratio has been found to correlate well with its CSF levels as well as with amyloid positron emission tomography (PET) findings, and its reduction is associated with cognitive decline and the risk of progression to dementia in cognitively unimpaired individuals, people with subjective cognitive impairment (SCI), and mild cognitive impairment (MCI) patients [12–16].

Plasma pTau181 has been emerging as an even more specific and sensitive biomarker for AD [17–19]. It shows a good correlation with its levels in the CSF [17], and has been associated with both amyloid and tau PET positivity [17,19]. It is estimated that plasma pTau181 can detect pathological accumulation of tau approximately 6 years earlier than tau PET [20]. Increased levels of plasma pTau181 are found in amyloid-positive cognitively unimpaired individuals, and pTau181 increases further in amyloid-positive MCI and AD patients, while it is not increased in several other neurodegenerative diseases and clinical mimics of AD [17,20].

Plasma glial fibrillary acidic protein (GFAP), which is a marker of neuroinflammation and reflects astrocytosis, is also considered to be associated with amyloid deposition in healthy controls, SCI, MCI and AD dementia patients [21–24]. This can be observed even in cognitively normal individuals with a normal amyloid status [23], and some evidence suggests that it is better than Aβ42/40 ratio in predicting the positivity of amyloid PET [21,24]. However, being a marker of neuroinflammation, raised GFAP levels are not specific to AD, and are also increased in many other neurological diseases [25]. Similarly to GFAP, another plasma biomarker that can be altered across several neurological disorders is neurofilament light chain (NfL). High baseline levels of NfL, an index of the rate of axonal injury, are strongly linked to markers of

neurodegeneration such as CSF total tau (t-tau), magnetic resonance imaging (MRI) atrophy and FDG-PET hypometabolism [26–28].

Some studies have attempted to examine the relationship between plasma biomarkers and cognitive performance, but to the best of our knowledge, the cognitive tests used were not digital online measures. For example, the plasma Aβ42/40 ratio is associated with scores on the Face Name Associative Memory Exam, cross-sectionally in SCI patients [29]. Performance on this test is also significantly correlated with amyloid burden measured by amyloid PET in cognitively unimpaired individuals [30]. Baseline levels and longitudinal increases in plasma pTau181 are associated with a decline on standard tests of cognition such as the mini mental state examination (MMSE), as well as with amyloid and tau accumulation and brain atrophy [18,31–33]. Another pTau isoform, plasma pTau217, was found to increase longitudinally in cognitively unimpaired people and MCI patients with evidence of amyloid positivity (A+), and in MCI patients who converted to AD over 6 years [34]. In the same study, an increase in pTau217 over time was correlated with worsening cognition on the MMSE and modified Preclinical Alzheimer's Cognitive Composite (mPACC) [35] in cognitively unimpaired and MCI participants.

GFAP has been linked to cognitive performance on standard tests of cognition in plasma and serum cross-sectionally [24,36], to a decline in cognition over time [37], and to clinical conversion to AD [22]. High baseline levels of NfL are also correlated with greater cognitive impairment trans-diagnostically [38], and a longitudinal increase in NfL is associated with worse cognitive scores in cognitively unimpaired A+ individuals [26] and patients with evidence of cognitive deficits (such as MCI and AD) [28,39]. These changes, however, do not seem to be specific to AD [40].

A recent report compared the head-to-head performance of different plasma biomarkers in cognitively unimpaired A+ individuals [41]. pTau217, pTau181, and GFAP were associated with cognitive decline (on the mPACC and MMSE), while pTau 231 and NfL did now show such an association [41]. The authors concluded that pTau217 alone was the strongest individual predictor of cognitive impairment. This was also observed in the whole sample of cognitively unimpaired individuals (whether A+ or A-). However, standard neuropsychological tests such as the mPACC and the MMSE still require a dedicated face-to-face appointment, which limits their use for large-scale population screening.

Here, we investigated whether plasma biomarkers (Aβ42/40 ratio, pTau181, GFAP, and NfL) are associated with several digital online cognitive metrics, measuring visual short-term memory (VSTM), long-term memory (LTM), visuospatial copying, executive function, and processing speed in a cohort of AD patients and two samples of elderly healthy controls, one of which also underwent blood collection for plasma biomarker measurement.

Methods

Ethics

Ethical approval was granted by the University of Oxford ethics committee (IRAS ID: 248379, Ethics Approval Reference: 18/SC/0448). All participants gave written informed consent prior to the start of the study.

Participants

53 EHC and 46 AD patients were recruited from the Cognitive Disorders Clinic at the John Radcliffe Hospital in Oxford. Alzheimer's disease dementia patients were defined as having Alzheimer's disease clinical syndrome according to the 2018 criteria[42] as they did not have ATN confirmation prior to enrollment in the study and will be subsequently referred to as AD. Elderly healthy controls were > 50 years old, had no psychiatric or neurological illness, were not on regular psychoactive drugs, and all scored above the cut-off for normality (88/100 total ACE score). The groups were not statistically different in age, gender or education level (Table 1).

Participants underwent blood collection, face-to-face standard cognitive, and online remote digital cognitive testing, and self-reported motivation and mood indices were collected (see Figure 1 for study flow).

To get a normative dataset for the digital cognitive battery, we recruited 256 participants above 50 years old through the Prolific participant recruitment platform (prolific.co). All participants were residents of the UK, had English as their first language and self-reported to be neurologically healthy. Four participants were excluded because they failed attention checks during the testing. All participants had normal or corrected-to-normal vision acuity and no colour blindness.

Measurements of plasma biomarkers: Aβ42, Aβ40, pTau181, NfL and GFAP

Four plasma biomarkers were assayed:

- Amyloid pathology ('A'): Aβ42, 40 and the Aβ42/40 ratio, which is a better measure of amyloid pathology than Aβ42 alone [12,43].
- Tau pathology ('T'): pTau181, which is a specific and sensitive marker of tau pathology in the blood and is highly predictive of tau PET positivity [18].
- Neurodegeneration ('N'): NfL, the most commonly used blood-based biomarker reflecting the rate of neurodegeneration occurring in the brain [44].
- Astrocytosis: GFAP, an established marker of astrocytosis and synaptic plasticity [23].

Supplementary Figure 1 presents the outline of plasma biomarker protocol. Blood was collected in 6 ethylenediaminetetraacetic acid (EDTA) tubes (10 ml each), and centrifuged (1800 g, RT, 10 minutes). The EDTA tubes were filled completely and gently inverted after

collection to avoid coagulation. After centrifugation, plasma from all 6 tubes was transferred into one 50-mL polypropylene tube, mixed, aliquoted into 0.5 mL polypropylene tubes (Fluid X, Tri-coded Tube, Azenta Life Sciences), and stored at 4°C, until (< 8 hours) it was transferred into a -80°C freezer. The time between blood collection and centrifugation was < 30 minutes. Transfer time between 4°C and -80°C storage was < 20 minutes, and the samples were kept refrigerated during transport. All cryovials were anonymized, and the unique cryovial code was logged into a secure database, linked to the participant's anonymous code and visit number. A separate paper-based document (i.e., a sample log) was filled in at collection, with the unique participant's anonymized code, gender, date of birth, hospital number, date, and time of different pre-processing steps (collection, start and end time of centrifugation, time into the 4°C fridge, and -80°C storage).

Samples were shipped in dry ice to the Biomarker Factory / Fluid Biomarker Laboratory, UK Dementia Research Institute at University College London (UCL), London. The Dementia Research Institute (DRI) laboratory staff carried out the analyses. Plasma Aβ40, Aβ42, GFAP, and NfL were measured by Single molecule array (Simoa) technology using the Neurology 4-plexE assay on an HD-X analyser (Quanterix), according to manufacturer's instructions. Plasma pTau181 was also measured by Simoa using the pTau-181 Advantage assay on an HD-X analyser (Quanterix). Briefly, samples were thawed at 21°C, and centrifuged at 10,000 RCF for five minutes at 21°C. On-board the instrument, samples were diluted 1:4 with sample diluent and bound to paramagnetic beads coated with a capture antibody specific for human Aβ40, Aβ42, GFAP, NfL and pTau-181. Aβ40-, Aβ42-, GFAP-, NfL- and pTau181-bound beads were then incubated with a biotinylated anti-Aβ40, anti-Aβ42, anti-GFAP, anti-NfL and anti-pTau181 detection antibodies in turn conjugated to streptavidin-β-galactosidase complex. Subsequent hydrolysis reaction with a resorufin β-D-galactopyranoside substrate produces a fluorescent signal proportional to the concentration of Aβ40, Aβ42, GFAP, NfL and pTau181 present. Singlicate measurements were taken of each sample. Sample concentrations were extrapolated from a standard curve, fitted using a 4-parameter logistic algorithm. Intra-assay and inter-assay CVs were less than 10% and 15% respectively, as determined by 8 quality controls following the same principles.

Face-to-face neuropsychological screening

Following plasma collection, all participants completed the Addenbrooke's Cognitive Examination-III (ACE) [45] in person at the time of the visit as a standard test of cognition.

Digital cognitive test battery: Oxford Cognition Online Platform

Participants also completed a sequence of computerised cognitive tasks from OCTAL (Oxford Cognitive Testing Portal, available at https://octalportal.com) (Figure 2). The tasks include Oxford Memory Test (OMT), Object-in-Scene Memory Task (OIS), drag-and-drop version of Rey-Osterrieth Complex Figure (ROCF), Trail Making Test (TMT), Digit Symbol Substitution Test (DSST) and Freestyle Corsi Block Task (CORSI) (Figure 2). They measure distinct aspects of human cognition, various forms of memory, attention, and executive functions. They were adapted from established behavioural paradigms or neuropsychological tests, whilst being robust against the type of device that a person is tested on. These six tasks can be tried at https://octalportal.com. The tasks were conceived and designed by S.Z. and M.H.. Most of the tasks were built using the PsychoPy Builder (PsychoJS, version 2022.2.4) with custom-written codes in Javascript, with one exception: the Rey–Osterrieth Complex Figure

(see details below), which was fully custom-written in HTML5 with JavaScript. All tasks were hosted on the server system Pavlovia.org.

A link with a unique patient and visit identifier was sent to the participants' email address the same day as the in-person visit when blood was collected. Participants were required to use Chrome or Firefox browsers on a desktop, laptop or tablet with any operating system. They were encouraged to complete the online tests within a week maximum. After two weeks, the completion of the online tests was reviewed, and participants who did not complete the tasks within that time frame were prompted via email to do so. Tests completed more than 3 weeks after the blood samples were discarded.

Since behavioural science is increasingly acknowledging that human behaviour sampled for convenience only across university populations may be WEIRD (Western, Educated, Industrialised, Rich and Democratic) [46], the healthy controls recruited around the university (EHC1) may not necessarily be representative of the general population. Thus, we recruited 256 healthy elderly participants online through Prolific.co as a normative group (group EHC 2, see Table 1 for basic demographics). To account for the effect of age on cognitive metrics, the cognitive performance of all participants was then transformed into z-score based on EHC 2 group (see details below "Normalisation for digital cognitive metrics"; Table 2 and Supplementary Table 1).

Oxford Memory Task (OMT)

OMT is the "What was where?" visual short-term memory experiment, which has been described in previous publications [4,47,48]. In this remote online version, participants were presented with one or three fractal patterns positioned at various locations on-screen for 3 seconds (Figure 2). Then, a 4-second delay was accompanied by a black screen. Subsequently, one of these fractal patterns was shown alongside a foil pattern. The two patterns were shown along the vertical meridian with 4 cm separation, with the order of the target and foil randomised across trials. Participants must identify which pattern they just saw (identification performance) by clicking the target pattern and drag it to its proper location on the screen (localisation performance). The foil was not a totally novel pattern; rather, it was part of the general pool of fractal images presented across the experiment. But the exact colour and shape of the foil never showed up as one of the patterns to remember.

Each participant performed a practice block of 6 trials including 3 trials with 1 item followed by 3 trials with 3 items. This is followed by a main test block of 40 trials, including 20 trials of 1 item and 20 trials of 3 items. The order of trials was randomised online. No feedback was given during practice or main test blocks. Fractal stimuli were drawn from a library of 196 pictures of fractals (http://sprott.physics.wisc.edu/fractals.htm), including 49 different shapes and each shape containing 4 colour variations.

As participants did the task remotely with their own devices, to ensure that the size of stimuli was physically the same across different devices, a card calibration procedure, previously described and validated [49], was employed prior to certain tasks. Participants are instructed to place a bank card or card of comparable size on the screen, and adjust the slider until the size of the image of the card on the screen matches the size of the physical card. This allows us to estimate screen distance by calculating the display's logical pixel density in pixels per centimetre. After successful calibration, the diameter of the fractal stimulus is 2 cm. A Matlab

script (MathWorks, Inc.) was used to determine the fractals' locations in a pseudorandom manner with a few constraints. In order to avoid crowding and create a clear zone around the items' original locations, which is essential for the analysis of localization errors, fractals were never placed closer than 3 cm to one another.

Eight cognitive metrics were extracted from this task: Identification Accuracy (proportion of correct object identification), Location Error (distance between response and target), Identification Time (reaction time to identify target), Localisation Time (reaction time to place object), Target Detection (rate of detecting correct object and placed at target location, see Figure 2), misbinding (rate of placing target at a non-target location), guessing (rate of placing target randomly) and imprecision (how precise spatially is the response).

Object-in-Scene Memory Task (OIS)

This test provides measures of identification accuracy, precision of spatial localization and semantic accuracy in visual STM and LTM (Figure 2). Participants were presented with a photo of an everyday scene and instructed to remember a particular object placed in the picture. To aid effective encoding, the participant was also asked to click on the displayed object. Subsequently, 20 different objects were presented, and the participant was asked to choose the correct object and place it in the remembered location in the scene. To ensure that they were not simply remembering the semantic information or name of the object, the object pool contained a foil that matches the target's category (e.g., two guitars of different colour and shape). After 5 different object and scene pairs, participants were asked to reproduce the object-scene associations probed (delayed recall, after 3-4 minutes). There were a total of 20 trials divided into four blocks, and the order of the pairs was randomised.

Three metrics were extracted from the task for both immediate and delayed recall stages: object identification accuracy (proportion of trials in which participants correctly identified the original object; chance level = 5%), semantic identification accuracy (proportion of objects correctly recalled as belonging to the same semantic category as the target; chancel level = 10%), location error (the distance from the original target item location to the centre of participant's response location; centimetre as unit).

Rey-Osterrieth Complex Figure (ROCF)

This task is a digitised version of the traditional pen-and-paper test[50]i . which is an established measure of visuospatial abilities (Figure 2). The original ROCF task requires the participant to draw a complex line-drawing freehand, first by replicating an existing figure (copy), and then again from memory (immediate recall). Our digitised version does not require hand drawing. Instead, the figure is split into 13 independent elements, and participants are required to drag each element into an empty canvas to copy the figure. Each test was automatically scored using an offline MATLAB-based algorithm. In contrast to the discrete score used in the pen-and-paper version, the score of our digital version provides a continuous measure of precision. The middle large rectangle is selected as the anchor point as a reference element. If the element is not placed (not present on the canvas), there will be no score; otherwise, the distance from the large rectangle is computed. As a measure of imprecision for each element, the absolute difference between the ideal distance and the actual distance is then calculated. The absolute error is then scaled using a logarithmic function: if the element is placed relatively

correctly, the difference in the distance from the big rectangle is computed; if the element is placed too far, the score is zero. The normalised absolute difference is then subtracted from 1 to calculate the score for this element. The sum of all element scores is 13, but the results are scaled to a percentage to match the original 36 element picture. Our version's scoring is consistent with the pen-and-paper scoring guide, as the participant receives one point for correctly positioning the element and no score if the element is placed incorrectly or not at all on the canvas. This task and scoring have been validated with the in-person traditional 36 items pen-and-paper test and manual scoring with standard scoring guide [57] in healthy participants before the start of the study. The percentages obtained at the copy and immediate recall – ROCF copy and recall scores – were used as metrics of interest.

Digit Symbol Substitution Test (DSST)

DSST provides a measure of processing speed. In this digitised version, participants were required to match symbols to digits according to a key located at the top of the page (Figure 2). The key consisted of 9 symbols next to the digits 1–9. At the bottom of the screen, there was a row of 9 randomised symbols. Participants reported the digit that corresponded to each symbol by clicking on the correct digit. The row was refreshed once all 9 were answered. Participants were allowed two minutes to answer as many as they could and the number of correct matches within the allowed time (DSST - correct responses).

Trail Making Test (TMT)

TMT is a standard test of processing speed and executive functions [51,52]. In this online version (Figure 2), 25 circled numbers are presented on-screen, and participants are instructed to connect them by clicking the circles in order as fast as possible. It contains three trials of Task A (where the order is 1-2-3-4-5-6-…) and three trials of Task B (order 1-A-2-B-3-C-…). Each participant sees six different trail maps randomly chosen from a pool of 100 pre-made maps, generated using a “divide-and-combine” approach [53]. The task also included a control condition of four trials to assess basic motor speed, where participants are presented with two circles located at two opposite corners of the screen. One is labelled with 1 and the other with 2, and participants were instructed to connect 1 with 2 as quickly as possible. The average of the reaction times of the TMT was used as a variable of interest.

Freestyle Corsi Block Task (CORSI)

This task is a modification of the Corsi Block Tapping Task [54], which is a standard measure of visual STM. In the original version, participants were presented with a set of nine identical wooden blocks positioned on a board. Subjects were required to point at the blocks in the order they were presented. They started with sequences of smaller blocks, and the number of blocks increased during the test. In the most computerised version of the task, the participant is shown several identical blocks that are in fixed locations spread across the screen [55]. Blocks then light up in sequence and the participant must remember which blocks lit up and in what order. In this digital version, the blocks’ locations were not fixed (‘Freestyle Corsi Block Task’). In an n-location trial, a 1-cm-wide red dot appeared at a random location on the screen, disappeared after 1 second, and reappeared at another random location on the screen (for n >1), and this process is repeated n times up to a sequence of 5 items. Once the sequence

has finished, after a 1 second break the participant could freely click anywhere on-screen to indicate where each dot appeared in sequence. The location error was calculated as the average distance between the response and the target location. The task was divided into five blocks, each block having five trials of a n-location sequence (i.e., 5 blocks of 1-item, 5 blocks of 2 items, up to 5 blocks of 5 items). The average of the reaction times of the 5 conditions was chosen as a variable of interest.

Questionnaire-derived motivation and mood metrics

All participants also completed two questionnaires which were hosted on Qualtrics: (1) Apathy Motivation Index (AMI), an 18-item questionnaire, sub-divided into three subscales of apathy: emotional, behavioural and social apathy [56] and (2) Geriatric Depression Scale (GDS), short form, which includes 15 questions. It is a screening tool designed to assess depressive symptoms in elderly people. A total score greater than 5 indicates probable depression [57]. Within each questionnaire, a validation question was embedded: "This is a validation question. Please choose 'No'."

Statistical analysis

For analysis and data visualisation purposes MATLAB (version R2023a), R studio (version 12.0), JASP (version 0.16.4) and SPSS (version 29.0) were used. Demographics, cognitive tests and plasma biomarkers levels were compared using a Mann-Whitney U test for continuous variables, while χ^2 test was used for categorical variables such as gender. P-values were two-tailed with statistical significance set at $p<0.05$ for all analyses. Rank-biserial correlation was used for effect size estimate. If data from multiple visits was available, averaged values per participant across visits were used.

Normalisation for digital cognitive measures

Z-score (i.e. number of standard deviations from the mean of the normative population in the similar age (± 3 yrs)) was computed for each variable and each subject, based on a normative population of 256 online participants above 50 years old (EHC 2, see Table 1 for demographics).

Correlation between digital cognitive metrics and plasma biomarkers

Plasma biomarkers' log10 transformed values and Z-scores of digital cognitive measures were used for correlation and linear regression analyses. Correlations between digital cognitive metrics and plasma biomarkers were assessed with Spearman's rank test, using age, sex, and education as covariates. The Benjamini–Hochberg method, which controls the False Discovery Rate (FDR), was used to correct for multiple comparisons.

Machine learning for group classification

Additionally, machine learning was applied to predict group classification and plasma biomarkers levels, firstly using MATLAB-based algorithms for feature ranking, to estimate the absolute contribution of each variable. The *fscchi2* function in MATLAB was used to predict group classification, while the *fsrftest* function was used to predict continuous variables, i.e., plasma biomarkers' levels. Rank's importance scores were then transformed into p-values by

calculating the exponential of the negative scores. Secondly, we applied the R-based MuMIn package to test which combinations of biomarkers would best predict group classification and plasma biomarkers levels. For predicting groups, we used logistic regression, while for predicting pTau181 level and Aβ42/40 ratio linear regression. The MuMIn package then uses the dredge function to achieve model selection, with the best performing model having the lowest corrected Akaike information criterion (AIC). ROC curves and AUCs were then computed for the model of interest. The pROC package in R with De Long's test was used to compare model performance in direct comparisons between two ROC curves.

Results

Participants tests and plasma biomarkers overview.

AD patients were significantly worse in all digital cognitive metrics with high effect sizes (|rrb| > 0.5) compared to matched controls (EHC1), see Figure 3 for distribution comparison for key cognitive metrics, and Supplementary Figures 2 and 3 for all metrics and online normative data. AD showed a large decline in executive functions, indexed by trail making test (TMT) and digit symbol substitution (DSST), as indicated by on average 8.5 standard deviation (SD) below expectation. Overall, AD patients were 2.0-7.5 SDs below expectation in both identification and localisation of recalling remembered items in working memory (OMT and Freestyle Corsi Block Task (CORSI)) and LTM (Object-in-Scene Memory Task (OIS) delayed recall). Noticeably, AD patients were particularly impaired at memory recall (OIS and Rey Osterrieth Complex Figure (ROCF)-recall). For example, both EHC groups could normally recall >90% of objects correctly with a very precise spatial memory (1 cm location error); in contrast, although AD remembered the object's semantic category (for example, it was a guitar), they could only recall 72.9% of the objects (but which guitar?) accurately with an average location error of 7.5 cm away from the centre of the object (which is 2 cm wide). Similarly, EHCs recalled 80% of ROCF immediately, but AD patients on average scored less than 50% (5.2 SD below expectation).

Compared to the online participants (EHC 2), EHC 1 performed slightly but significantly better in many cognitive metrics (see Supplementary Table 1). In our sample, this difference could not be explained by age, education level, or the testing environment (all completed remotely at home anonymously). Online participants in EHC2 performed particularly worse in the Oxford Memory Task (OMT), where they made significantly more misbinding errors and faster at localisation compared to the participants we tested locally. This group difference might be due to a speed-accuracy trade-off in EHC 2 group; in this online group, participants with shorter localisation time were associated with more misbinding errors (Pearson $r = -0.22$, $p=0.003$), while in contrast no correlation between speed and accuracy was found in EHC 1 ($r=-0.07$, $p = 0.64$).

Regarding plasma biomarkers, as expected, patients with AD had higher mean levels of pTau181, GFAP and NfL and lower Aβ42/40 ratio compared to controls (Table 1).

The standard neuropsychological test used, the Addenbrooke's Cognitive Examination (ACE), was significantly different between AD and EHC1, and had a high effect size in group comparisons, which was expected as it was the only test used for diagnosis. As expected, AD patients were in general more apathetic and depressed compared to EHCs (Table 1). Across

all participants, the apathy level overall strongly correlated with the level of depression (Pearson's r = 0.46, p<0.001), but it did not significantly correlate with any of our online cognitive metrics (Supplementary Figure 4).

Relationships between plasma biomarkers and cognitive metrics.

The relationships between all plasma biomarkers and digital cognitive metrics are visualised as a network plot in Figure 4a, in which the strength of the relationship is represented by the distance between the metrics. Among the four plasma biomarkers investigated in the present study, pTau181 was most strongly correlated with our digital cognitive metrics, which all clustered on the right of the plot. In contrast, Aβ42/40 ratio showed the weakest relationship with cognitive performance as well as with the other three plasma biomarkers. Among the digital cognitive metrics, OIS and the ROCF were the most closely associated with the plasma biomarkers. OMT and TMT were the tests that showed the weakest correlations with plasma biomarkers.

This pattern of relationships can also be appreciated when looking at the individual correlation between each pair of biomarkers and cognitive metrics (Figure 4b). Across the different tasks examined, multiple metrics of STM were correlated with pTau181, GFAP and NfL levels; the better the performance, the lower the levels of these three plasma biomarkers. Similarly, these plasma levels were also correlated with executive function metrics such as DSST and TMT, and with visuospatial ability as indicated by the ROCF copy score. In contrast, Aβ42/40 ratio was only weakly associated with STM metrics and was not associated with the LTM related metrics in OIS (e.g., delayed recall accuracy and localization error). However, Aβ42/40 ratio levels were associated with the immediate recall scores of ROCF. When looking closer at the responses in the working memory task OMT, we found that pTau181 was associated with multiple metrics, while GFAP was specifically related to misbinding and guessing rates; people with higher GFAP tended to make more localisation errors, such as placing the target object in a non-target location or in a random place. Additionally, in OMT NfL levels were associated with the imprecision of the response (i.e., the spatial spread of the responses around the target response, see more in Methods).

Which plasma/cognitive metric best predicts AD?

The selected variables were then ranked according to their importance in predicting group classification, i.e., AD or EHC (Figure 5a). The rank represents the negative log of the p-values. In this sample, the recall of the ROCF and the Identification accuracy of the OIS had a higher rank compared to pTau181, and all cognitive metrics ranked higher than all the other plasma biomarkers. All tests and biomarkers were significant predictors of the group (all $p < 0.001$ except Aβ42/40 ratio, $p= 0.002$).

We then explored which were the best predictors of tau and amyloid pathology, indexed respectively by pTau181 and the Aβ42/40 ratio. Most of our digital cognitive tests, except the CORSI, were better predictors of pTau181 levels compared to the other plasma biomarkers (Supplementary Figure 5a). The most predictive cognitive tests were the ROCF and TMT ($p < 0.001$), followed by OIS ($p = 0.003$), OMT ($p = 0.004$), DSST ($p = 0.004$), NfL ($p = 0.005$), Aβ42/40 ($p = 0.02$), while GFAP ($p = 0.06$) and CORSI ($p = 0.13$) were not statistically significant predictors of pTau181 levels. Conversely, pTau181 was the only statistically significant predictor of amyloid burden ($p = 0.02$) (Supplementary Figure 5b). The best

performing digital cognitive test to predict amyloid burden was the OIS task, but it was not statistically significant.

Model comparison using the MuMIn R function was then used to choose the best combination of plasma biomarkers and cognitive metrics in predicting group classification, pTau181 and the Aβ42/40 ratio levels. The best model for predicting groups consisted of pTau181, recall of the ROCF, and Immediate Object Accuracy of the OIS (Figure 5b and Supplementary Figure 5c), with an AUC of 1. In comparison, the recall of the ROCF alone had an AUC of 0.946, while pTau181 had an AUC of 0.911, which were, however, not statistically significantly different from the best model (ROCF: $Z = 1.6$, $p = 0.11$, pTau181: $Z = 1.9$, $p = 0.06$), nor between them ($Z = 0.60$, $p = 0.55$). When Lasso penalization was introduced to avoid perfect separation of the best model, it still performed significantly well (AUC = 0.92). In comparison, ACE had an AUC of 0.97 in discriminating between groups, which was, however, not different compared to the best performing digital metric, ROCF ($Z = 0.06$, p-value = 0.95).

The best model for predicting pTau181 levels consisted of Aβ42/40, NfL, OIS and ROCF, which had an adjusted R^2 of 0.44 (Supplementary Figure 5c). The winning model for predicting Aβ42/40 levels consisted of pTau181, CORSI, OIS, OMT and ROCF, but had an overall poorer model fit with an adjusted R^2 of 0.24 (Supplementary Figure 5c). If single metrics were evaluated, in predicting pTau181 levels, OIS and ROCF had an adjusted R^2 of, respectively, 0.33 and 0.25, while in comparison, ACE had an adjusted R^2 of 0.14. AICs were -18.57 (OIS), -9.90 (ROCF) and 8.43 (ACE), with the best-performing model being the one containing OIS (more negative).

Discussion

The findings reported here demonstrate that AD patients have impaired performance on our digital cognitive tests and that this is closely related to pathological blood-based biomarkers of the disease (Figures 3 and 4). Levels of plasma pTau181, GFAP and NfL were all highly correlated with several cognitive metrics, with pTau181 being the biomarker that showed the closest association to digital cognitive performance (Figure 4a). Overall, the results of this study show that it is feasible to deploy digital cognitive testing in AD, which would have the potential to make a significant impact on several fronts. Screening for the disease, recruitment and stratification into clinical trials, and longitudinal follow-up in intervention studies could all be transformed if cognitive testing were to be conducted robustly, remotely, and frequently [11]. Digital cognitive testing could make this happen.

In terms of single metrics, two measures of visual memory performed best. These were the recall on the ROCF test (a measure of visual episodic memory) and Immediate Object Accuracy on the OIS task (a measure of visual short-term memory). In fact, they were better predictors of group classification (AD or EHC) compared to all plasma biomarkers (Figure 5a). Further, performance on the TMT, CORSI, DSST, and OMT predicted groups better than all plasma biomarkers except pTau181. If used in combination, the best-performing model to predict the group was the one combining pTau181, recall of the ROCF and Immediate Object Accuracy of the OIS with an AUC of 1, achieving perfect separation of patients from controls (Figure 5b). The best-performing digital metric, recall of ROCF, performed equally well as pTau181 in group classification. Further, the model combining pTau181, the recall of ROCF

and Immediate Object Accuracy of OIS was not statistically significantly better than the model containing ROCF or pTau181 alone.

The very high AUCs of these metrics in predicting groups in this small, highly selected sample might partially explain the lack of positive contribution of adding cognition to plasma biomarkers. With accuracy being at ceiling, further evidence is needed to establish whether combining pTau181 to digital metrics might be beneficial in a larger dataset including different populations such as individuals with SCI or MCI. A bigger sample would also be required to assess the performance of these metrics in people with preclinical AD versus amyloid and tau-negative healthy controls. Therefore, whether the combination of digital cognitive metrics and plasma biomarkers can be useful to stratify which individuals in the preclinical or prodromal phase of AD might be at risk of developing AD dementia remains to be established. However, it is encouraging that, even with a relatively small sample size, these metrics show a good correlation with several plasma biomarkers, surviving multiple comparisons and corrections for age, gender, and education, which are major confounders in both plasma biomarkers and cognitive assessments [7,58].

Notably, while performance on our digital cognitive tests was tightly associated with plasma biomarker levels, it was not statistically significantly correlated with measures of apathy and depression (Supplementary Figure 4). This supports the proposal that these metrics are related to cognitive performance and do not merely measure willingness to engage in the online task, which is a potential confounder. This type of control is important as patients with AD are often reported to have higher rates of anxiety and depression compared to age-matched healthy controls, a finding that was also present in our sample (Table 1).

A key finding of this study is that digital cognitive metrics were more tightly correlated with pTau181 than the Aβ42/40 ratio (Figure 4 and Supplementary Figure 5). This is not entirely surprising, as amyloid burden has been shown to have a weaker association with cognition compared to tau [59,60]. A large meta-analysis that investigated different indices of amyloid positivity in cognitively unimpaired elderly adults without blood-based biomarkers showed that although episodic memory might be correlated with amyloid burden, global cognition and executive functions are not if assessed by amyloid PET [61]. In our digital platform, higher amyloid burden indexed by a lower Aβ42/40 ratio was not uniquely associated with LTM abilities but was correlated with performance across multiple tasks, measuring visual STM and LTM, executive function, and visuospatial function (Figure 4b). One possible explanation for the better performance of pTau181 compared to the Aβ42/40 ratio lies in the fact that pTau181 in blood correlates well with both amyloid and tau PET [17], and not only with amyloid burden. Moreover, unlike GFAP and NfL, pTau181 is AD-specific [18].

Digital platforms are emerging as potential screening and diagnostic tools for people at risk of developing AD [62]. Most studies using such platforms have focused on screening healthy individuals [10]. Moreover, biomarker validation on these digital platforms is mostly limited to one single marker, frequently amyloid PET [10]. Some brief digital screening tools have demonstrated promise in differentiating amyloid-positive and tau-positive MCI patients (as measured by amyloid and tau PET) from MCI without evidence of amyloid or tau accumulation, but they are not very good at separating healthy controls from people with MCI or prodromal AD [63]. A digital version of the Face Name Associative Memory Exam, measuring episodic memory, has also been found to correlate with CSF levels of pTau181 and the Aβ42/40 [64]. The performance of these tests compared to plasma biomarkers of AD is currently unknown.

One of the strengths of the current study is the inclusion of tests measuring different cognitive domains and the use of different plasma biomarkers, measuring not only amyloid and tau accumulation but also neuroinflammation and neurodegeneration. To our knowledge, no study so far has investigated the relationship between these four biomarkers and performance on a digital platform in a mixed population of elderly healthy controls and AD patients.

The patient population included in this study was already at the AD dementia stage, where cognitive impairment is overt. In this sample, plasma pTau181 was the biomarker that was more closely associated with cognitive metrics and the best predictor of group classification. However, we cannot exclude that other biomarkers such as pTau217 could show an even higher association with cross-sectional or longitudinal cognitive function in the same population, as some evidence suggests [65,66]. Also, early markers of amyloid deposition such as pTau217, pTau213 and GFAP may be more closely linked with cognitive changes in the early phases of the disease [21,22,65,66].

Importantly, amongst the digital metrics, ROCF had comparable performance to the standard cognitive scores used (ACE) in group classification, and OIS and ROCF had better performance (higher adjusted R^2 and lower AIC) than ACE in predicting pTau181 levels. This is encouraging, as in the future, the combination of these measures might be used as a proxy for standard cognitive metrics while saving a considerable amount of time in face-to-face appointments.

To conclude, digital cognitive metrics were tightly associated with several plasma biomarkers, particularly pTau181, but also with GFAP and NfL, and to a much lesser extent with the Aβ42/40 ratio. Adding these metrics to pTau181 did not improve group classification in this sample, but the best performing metric, the recall of ROCF, performed at par with pTau181 levels. As plasma biomarkers are being proposed as equivalent to CSF biomarkers in the forthcoming NIA-AA revised criteria for AD [67], and given their increased use in clinical practice [68], implementation of a digital cognitive platform that has been validated with AD plasma biomarkers provides an important step forward for future large-scale deployment.

Data availability

De-identified data supporting this study may be shared based on reasonable written requests to the corresponding author. Access to de-identified data will require a Data Access Agreement and IRB clearance, which will be considered by the institutions who provided the data for this research.

Code Availability

The source code will be shared using a Creative Commons NC-ND 4.0 international licence upon reasonable written request to the corresponding author and requires a research use agreement.

Competing interests

H.Z. has served at scientific advisory boards and/or as a consultant for Abbvie, Acumen, Alector, Alzinova, ALZPath, Annexon, Apellis, Artery Therapeutics, AZTherapies, Cognito

Therapeutics, CogRx, Denali, Eisai, Nervgen, Novo Nordisk, Optoceutics, Passage Bio, Pinteon Therapeutics, Prothena, Red Abbey Labs, reMYND, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, Biogen, and Roche, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program (outside submitted work). The other authors declare no financial or non-financial competing interests.

Acknowledgement

This research was supported by funding from the Wellcome Trust and NIHR Oxford Health Biomedical Research Centre. S.T., S.Z. and M.H. were funded by the Wellcome Trust (206330/Z/17/Z). The project was further supported by a Guarantors of Brain post-doctoral fellowship to S.T.. H.Z. is a Wallenberg Scholar supported by grants from the Swedish Research Council (#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, and #ADSF-21-831377-C), the Bluefield Project, the Olav Thon Foundation, the Erling-Persson Family Foundation, 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, and the UK Dementia Research Institute at UCL (UKDRI-1003). The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Authors contribution

STo: conceptualization, participants' recruitment, project administration, resources management, data collection and curation, analysis, data visualisation, writing, SZ: conceptualization, software development, code development, digital platform data curation and management, analysis, data visualisation, writing, AS: cognitive testing data collection and curation, BA: biomarker samples preprocessing and data curation, AG: cognitive testing data collection and curation, AH: biomarker samples analysis, STh and SM: participants' recruitment, HZ: biomarker samples analysis, MH: conceptualization, supervision, resources management, funding acquisition, participants' recruitment, writing. All authors read and approved the final manuscript.

References

https://doi.org/10.1108/IJCS-12-2016-0002

1. Sims, J. R. *et al.* Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA* (2023)

doi:10.1001/jama.2023.13239.

2. van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine* **388**, 9–21 (2023).
3. Jack, C. R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119–128 (2010).
4. Liang, Y. *et al.* Visual short-term memory binding deficit in familial Alzheimer's disease. *Cortex* **78**, 150–164 (2016).
5. Bangen, K. J. *et al.* Elevated plasma neurofilament light predicts a faster rate of cognitive decline over 5 years in participants with objectively-defined subtle cognitive decline and MCI. *Alzheimer's & Dementia* **17**, 1756–1762 (2021).
6. Meier, I. B. *et al.* Using a Digital Neuro Signature to measure longitudinal individual-level change in Alzheimer's disease: the Altoida large cohort study. *npj Digit. Med.* **4**, 1–9 (2021).
7. Berron, D. *et al.* Feasibility of Digital Memory Assessments in an Unsupervised and Remote Study Setting. *Front. Digit. Health* **4**, 892997 (2022).
8. Brooker, H. *et al.* FLAME: A computerized neuropsychological composite for trials in early dementia. *Alzheimers Dement (Amst)* **12**, e12098 (2020).
9. Kochan, N. A. *et al.* Reliability, Validity, and User-Experience of Remote Unsupervised Computerized Neuropsychological Assessments in Community-Living 55- to 75-Year-Olds. *J Alzheimers Dis* **90**, 1629–1645 (2022).
10. Öhman, F., Hassenstab, J., Berron, D., Schöll, M. & Papp, K. V. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alz & Dem Diag Ass & Dis Mo* **13**, (2021).
11. Jutten, R. J. *et al.* Why a clinical trial is as good as its outcome measure: A framework for the selection and use of cognitive outcome measures for clinical trials of Alzheimer's disease. *Alzheimers Dement* **19**, 708–720 (2023).
12. Doecke, J. D. *et al.* Total Aβ42/Aβ40 ratio in plasma predicts amyloid-PET status, independent of clinical AD diagnosis. *Neurology* **94**, e1580–e1591 (2020).

13. Giudici, K. V. *et al.* Assessment of Plasma Amyloid-β42/40 and Cognitive Decline Among Community-Dwelling Older Adults. *JAMA Network Open* **3**, e2028634 (2020).
14. Stockmann, J. *et al.* Amyloid-β misfolding as a plasma biomarker indicates risk for future clinical Alzheimer's disease in individuals with subjective cognitive decline. *Alzheimer's Research & Therapy* **12**, 169 (2020).
15. Verberk, I. M. W. *et al.* Plasma amyloid is associated with the rate of cognitive decline in cognitively normal elderly: the SCIENCe project. *Neurobiol Aging* **89**, 99–107 (2020).
16. Verberk, I. M. W. *et al.* Plasma Amyloid as Prescreener for the Earliest Alzheimer Pathological Changes. *Annals of Neurology* **84**, 648–658 (2018).
17. 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).
18. 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).
19. 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).
20. Moscoso, A. *et al.* Time course of phosphorylated-tau181 in blood across the Alzheimer's disease spectrum. *Brain* **144**, 325–339 (2021).
21. Chatterjee, P. *et al.* Plasma glial fibrillary acidic protein is elevated in cognitively normal older adults at risk of Alzheimer's disease. *Transl Psychiatry* **11**, 1–10 (2021).
22. Cicognola, C. *et al.* Plasma glial fibrillary acidic protein detects Alzheimer pathology and predicts future conversion to Alzheimer dementia in patients with mild cognitive impairment. *Alzheimer's Research & Therapy* **13**, 68 (2021).
23. Pereira, J. B. *et al.* Plasma GFAP is an early marker of amyloid-β but not tau pathology in Alzheimer's disease. *Brain* **144**, 3505–3516 (2021).
24. Verberk, I. M. W. *et al.* Combination of plasma amyloid beta(1-42/1-40) and glial fibrillary

acidic protein strongly associates with cerebral amyloid pathology. *Alzheimer's Research & Therapy* **12**, 118 (2020).

25. Abdelhak, A. *et al.* Blood GFAP as an emerging biomarker in brain and spinal cord disorders. *Nat Rev Neurol* **18**, 158–172 (2022).
26. Benedet, A. L. *et al.* Plasma neurofilament light associates with Alzheimer's disease metabolic decline in amyloid-positive individuals. *Alzheimers Dement (Amst)* **11**, 679–689 (2019).
27. Marks, J. D. *et al.* Comparison of plasma neurofilament light and total tau as neurodegeneration markers: associations with cognitive and neuroimaging outcomes. *Alzheimer's Research & Therapy* **13**, 199 (2021).
28. Mattsson, N., Cullen, N. C., Andreasson, U., Zetterberg, H. & Blennow, K. Association Between Longitudinal Plasma Neurofilament Light and Neurodegeneration in Patients With Alzheimer Disease. *JAMA Neurology* **76**, 791–799 (2019).
29. Pascual-Lucas, M. *et al.* Clinical performance of an antibody-free assay for plasma Aβ42/Aβ40 to detect early alterations of Alzheimer's disease in individuals with subjective cognitive decline. *Alzheimer's Research & Therapy* **15**, 2 (2023).
30. Rentz, D. M. *et al.* Face-name associative memory performance is related to amyloid burden in normal elderly. *Neuropsychologia* **49**, 2776–2783 (2011).
31. Karikari, T. K. *et al.* Blood phospho-tau in Alzheimer disease: analysis, interpretation, and clinical utility. *Nat Rev Neurol* **18**, 400–418 (2022).
32. Mielke, M. M. *et al.* Comparison of Plasma Phosphorylated Tau Species With Amyloid and Tau Positron Emission Tomography, Neurodegeneration, Vascular Pathology, and Cognitive Outcomes. *JAMA Neurol* **78**, 1108–1117 (2021).
33. O'Connor, A. *et al.* Plasma phospho-tau181 in presymptomatic and symptomatic familial Alzheimer's disease: a longitudinal cohort study. *Mol Psychiatry* **26**, 5967–5976 (2021).
34. Mattsson-Carlgren, N. *et al.* Longitudinal plasma p-tau217 is increased in early stages of Alzheimer's disease. *Brain* **143**, 3234–3241 (2020).
35. Donohue, M. C. *et al.* The preclinical Alzheimer cognitive composite: measuring amyloid-

related decline. *JAMA Neurol* **71**, 961–970 (2014).

36. Oeckl, P. *et al.* Glial Fibrillary Acidic Protein in Serum is Increased in Alzheimer's Disease and Correlates with Cognitive Impairment. *J Alzheimers Dis* **67**, 481–488 (2019).
37. Rajan, K. B. *et al.* Remote Blood Biomarkers of Longitudinal Cognitive Outcomes in a Population Study. *Ann Neurol* **88**, 1065–1076 (2020).
38. Ashton, N. J. *et al.* A multicentre validation study of the diagnostic value of plasma neurofilament light. *Nat Commun* **12**, 3400 (2021).
39. Mielke, M. M. *et al.* Plasma and CSF neurofilament light: Relation to longitudinal neuroimaging and cognitive measures. *Neurology* **93**, e252–e260 (2019).
40. Moscoso, A. *et al.* Longitudinal Associations of Blood Phosphorylated Tau181 and Neurofilament Light Chain With Neurodegeneration in Alzheimer Disease. *JAMA Neurology* **78**, 396–406 (2021).
41. Mattsson-Carlgren, N. *et al.* Prediction of Longitudinal Cognitive Decline in Preclinical Alzheimer Disease Using Plasma Biomarkers. *JAMA Neurology* **80**, 360–369 (2023).
42. Jack Jr., C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia* **14**, 535–562 (2018).
43. Fandos, N. *et al.* Plasma amyloid β 42/40 ratios as biomarkers for amyloid β cerebral deposition in cognitively normal individuals. *Alzheimers Dement (Amst)* **8**, 179–187 (2017).
44. Ashton, N. J. *et al.* Increased plasma neurofilament light chain concentration correlates with severity of post-mortem neurofibrillary tangle pathology and neurodegeneration. *Acta Neuropathologica Communications* **7**, 5 (2019).
45. Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W. & Hodges, J. R. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* **55**, 1613–1620 (2000).
46. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav Brain Sci* **33**, 61–83; discussion 83-135 (2010).

47. Pavisic, I. M. *et al.* Visual short-term memory impairments in presymptomatic familial Alzheimer's disease: A longitudinal observational study. *Neuropsychologia* **162**, 108028 (2021).
48. Pertzov, Y. *et al.* Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. *Brain* **136**, 2474–2485 (2013).
49. Li, Q., Joo, S. J., Yeatman, J. D. & Reinecke, K. Controlling for Participants' Viewing Distance in Large-Scale, Psychophysical Online Experiments Using a Virtual Chinrest. *Sci Rep* **10**, 904 (2020).
50. Osterrieth, P. A. Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. [Test of copying a complex figure; contribution to the study of perception and memory.]. *Archives de Psychologie* **30**, 206–356 (1944).
51. Bowie, C. R. & Harvey, P. D. Administration and interpretation of the Trail Making Test. *Nat Protoc* **1**, 2277–2281 (2006).
52. Sherman, E., Tan, J. & Hrabok, and M. *A Compendium of Neuropsychological Tests: Fundamentals of Neuropsychological Assessment and Test Reviews for Clinical Practice*. (Oxford University Press, 2023).
53. Zeng, Z., Miao, C., Leung, C. & Shen, Z. Computerizing Trail Making Test for long-term cognitive self-assessment. *International Journal of Crowd Science* **1**, 83–99 (2017).
54. Corsi, P. M. Human memory and the medial temporal region of the brain. (McGill University, 1972).
55. Arce, T. & McMullen, K. The Corsi Block-Tapping Test: Evaluating methodological practices with an eye towards modern digital frameworks. *Computers in Human Behavior Reports* **4**, 100099 (2021).
56. Ang, Y.-S., Lockwood, P., Apps, M. A. J., Muhammed, K. & Husain, M. Distinct Subtypes of Apathy Revealed by the Apathy Motivation Index. *PLOS ONE* **12**, e0169938 (2017).
57. Yesavage, J. Geriatric depression scale. in *Psychopharmacology Bulletin* 709–711 (The Clearinghouse, 1988).

58. Ferreira, P. C. L. *et al.* Plasma biomarkers identify older adults at risk of Alzheimer's disease and related dementias in a real-world population-based cohort. *Alzheimer's & Dementia* **n/a**, (2023).
59. Giannakopoulos, P. *et al.* Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology* **60**, 1495–1500 (2003).
60. Ossenkoppele, R. *et al.* Accuracy of Tau Positron Emission Tomography as a Prognostic Marker in Preclinical and Prodromal Alzheimer Disease: A Head-to-Head Comparison Against Amyloid Positron Emission Tomography and Magnetic Resonance Imaging. *JAMA Neurology* **78**, 961–971 (2021).
61. Hedden, T., Oh, H., Younger, A. P. & Patel, T. A. Meta-analysis of amyloid-cognition relations in cognitively normal older adults. *Neurology* **80**, 1341–1348 (2013).
62. Staffaroni, A. M., Tsoy, E., Taylor, J., Boxer, A. L. & Possin, K. L. Digital Cognitive Assessments for Dementia. *Pract Neurol (Fort Wash Pa)* **2020**, 24–45 (2020).
63. Alden, E. C. *et al.* Diagnostic accuracy of the Cogstate Brief Battery for prevalent MCI and prodromal AD (MCI A+T+) in a population-based sample. *Alzheimer's & Dementia* **17**, 584–594 (2021).
64. Alegret, M. *et al.* Automatized FACEmemory® scoring is related to Alzheimer's disease phenotype and biomarkers in early-onset mild cognitive impairment: the BIOFACE cohort. *Alzheimer's Research & Therapy* **14**, 43 (2022).
65. Ashton, N. J. *et al.* Differential roles of Aβ42/40, p-tau231 and p-tau217 for Alzheimer's trial selection and disease monitoring. *Nat Med* **28**, 2555–2562 (2022).
66. Milà-Alomà, M. *et al.* Plasma p-tau231 and p-tau217 as state markers of amyloid-β pathology in preclinical Alzheimer's disease. *Nat Med* **28**, 1797–1801 (2022).
67. NIA-AA Workshop. NIA-AA Revised Clinical Criteria for Alzheimer's Disease. in (2023).
68. Hampel, H. *et al.* Blood-based biomarkers for Alzheimer's disease: Current state and future use in a transformed global healthcare landscape. *Neuron* S0896627323003902 (2023) doi:10.1016/j.neuron.2023.05.017.

Figures and Tables.

Figure 1. Study schematic.

AD = Alzheimer's disease. EHC = Elderly Healthy Control.

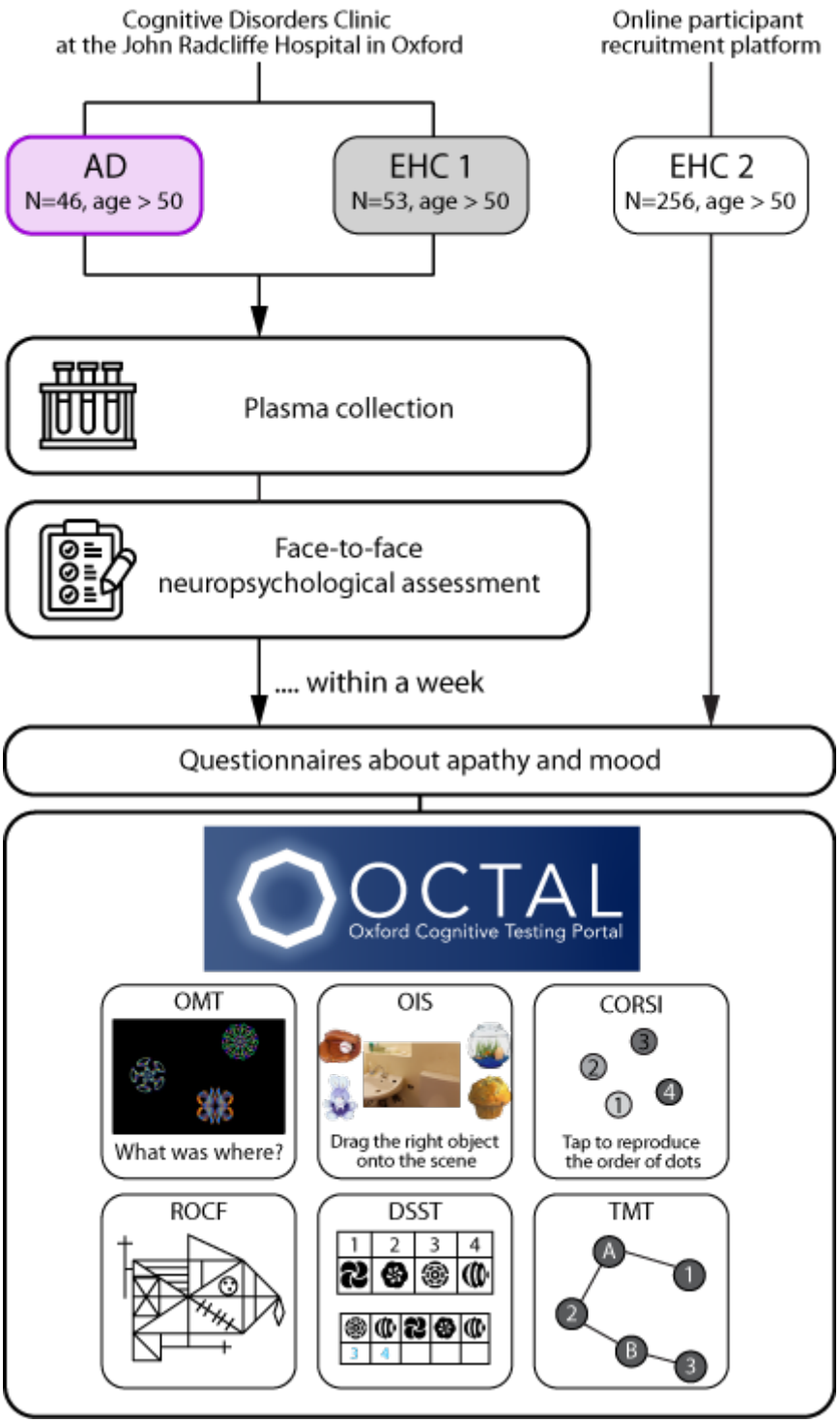

Figure 2. Experimental design of online digital cognitive tasks.

Oxford Memory Test (OMT) is a "What was where?" visual short-term memory experiment where participants are presented with one or three fractal patterns positioned at various locations on-screen for 3 seconds. After a 4-second delay, participants identify the fractal pattern shown before and move it to its remembered location. Thus, the response reflects how precisely the memory was recalled. Additionally, the Object-in-Scene task (OIS) measures long-term memory. Participants are shown with a photo of an everyday scene and asked to remember a particular object placed in the picture. Another memory task is Freestyle Corsi Block Task (CORSI), which involves remembering a sequence of random locations on the screen. Participants are then free to click anywhere on-screen to indicate the remembered sequence of locations. Participants in the Rey-Osterrieth Complex Figure (ROCF) test their visuospatial abilities by dragging 13 elements given to copy a complex line graph on an empty canvas. Digit Symbol Substitution Test (DSST) requires participants to match symbols to digits according to a key located at the top of the page. More correct matches were made in under two minutes indicating faster processing speed. Trail Making Test (TMT) is also an executive function task. Participants are instructed to connect 25 circled numbers by clicking the circles in order as fast as possible. This task contains three trials of Task A (where the order is 1-2-3-4-5-6-…) and three trials of Task B (order 1-A-2-B-3-C-…). Full versions of all tasks can be tried online at https://octalportal.com.

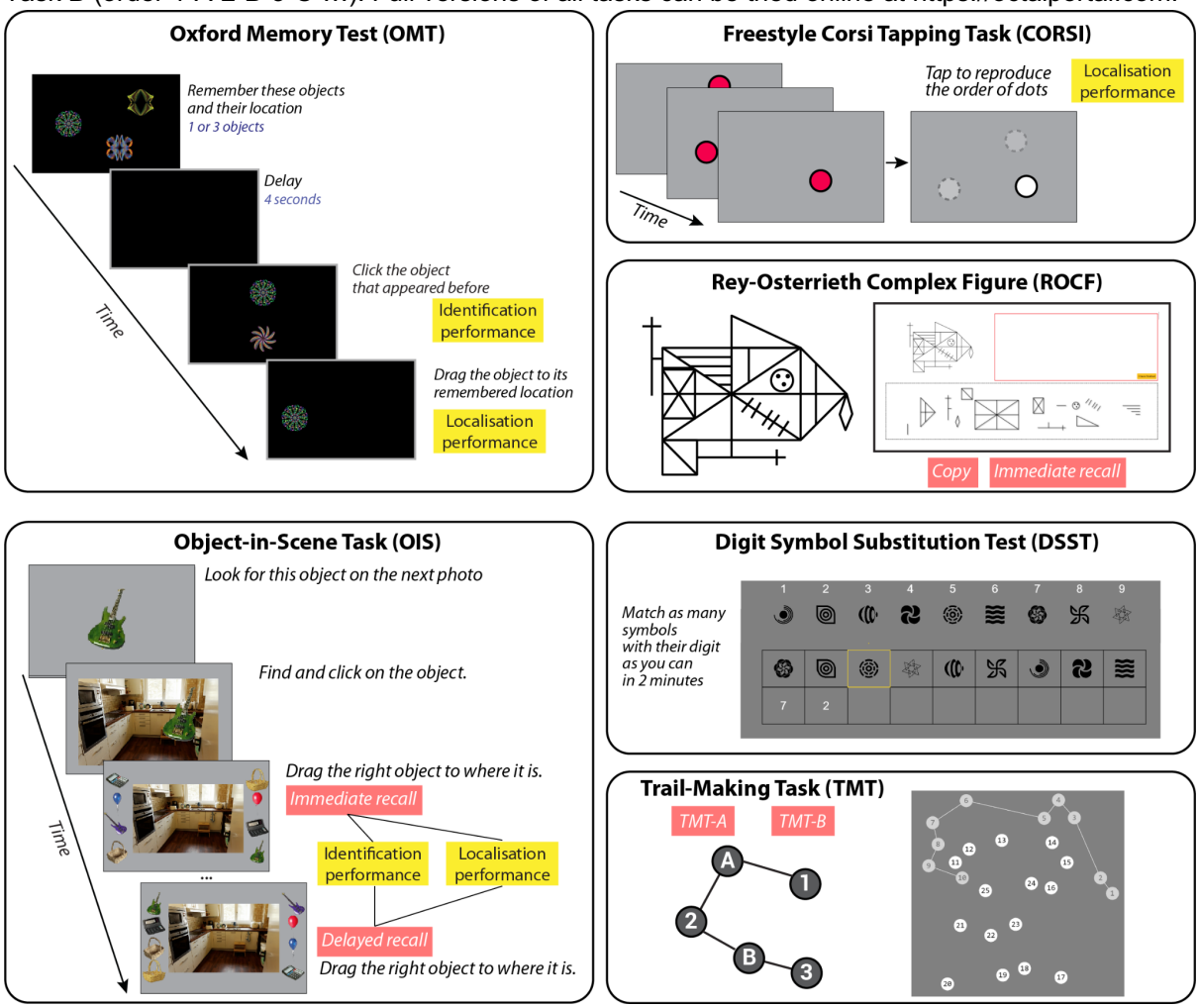

Figure 3. Digital cognitive metrics.

In all online tasks, AD patients (plotted in lilac) performed significantly worse and had higher variability compared to age-matched healthy controls (EHC1, plotted in grey). On the X-axis, cognitive performance is shown as z-scores derived from age-matched online normative data, except TMT for which the raw completion time is shown. Y-axis indicates the percentage of participants. See supplementary materials for the distribution plots for all other cognitive metrics.

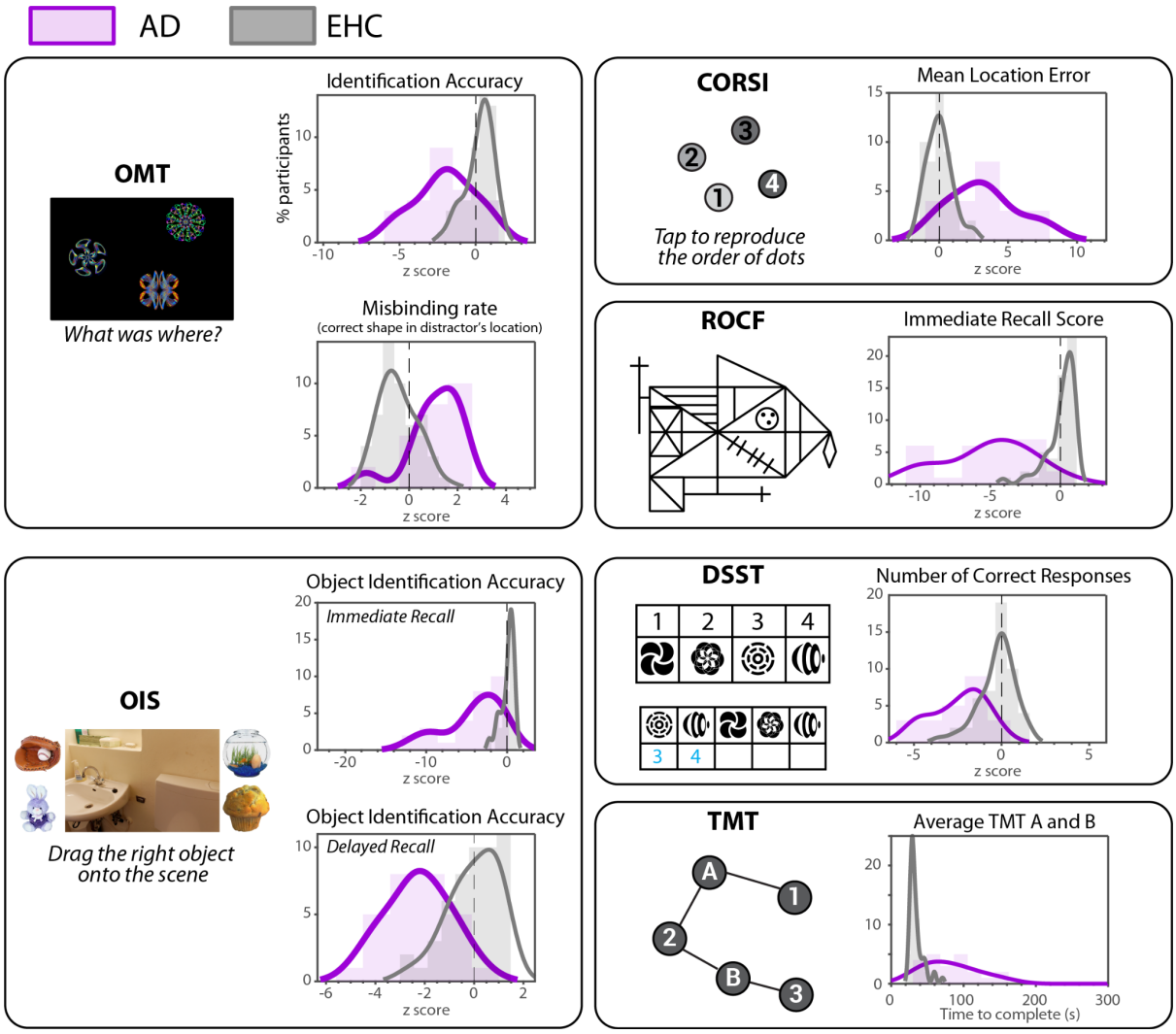

Figure 4. pTau181 shows the strongest relationship with digital cognitive tests.

(a) Network plot of relationships between all plasma biomarkers (clustered on the left) and online digital cognitive tasks (clustered on the right). Associations are presented in graded colours, where red is associated with a negative correlation and light blue with a positive correlation. The shorter distance between two metrics indicates a stronger relationship (larger correlation coefficient). For online tasks, only one metric was selected per task, according to the highest effect size in discriminating between groups. **(b)** Strength of the correlation is given by diameter of the circle, with positive correlations in blue and negative in red. All displayed correlations are significant after Benjamini–Hochberg correction for multiple comparisons.

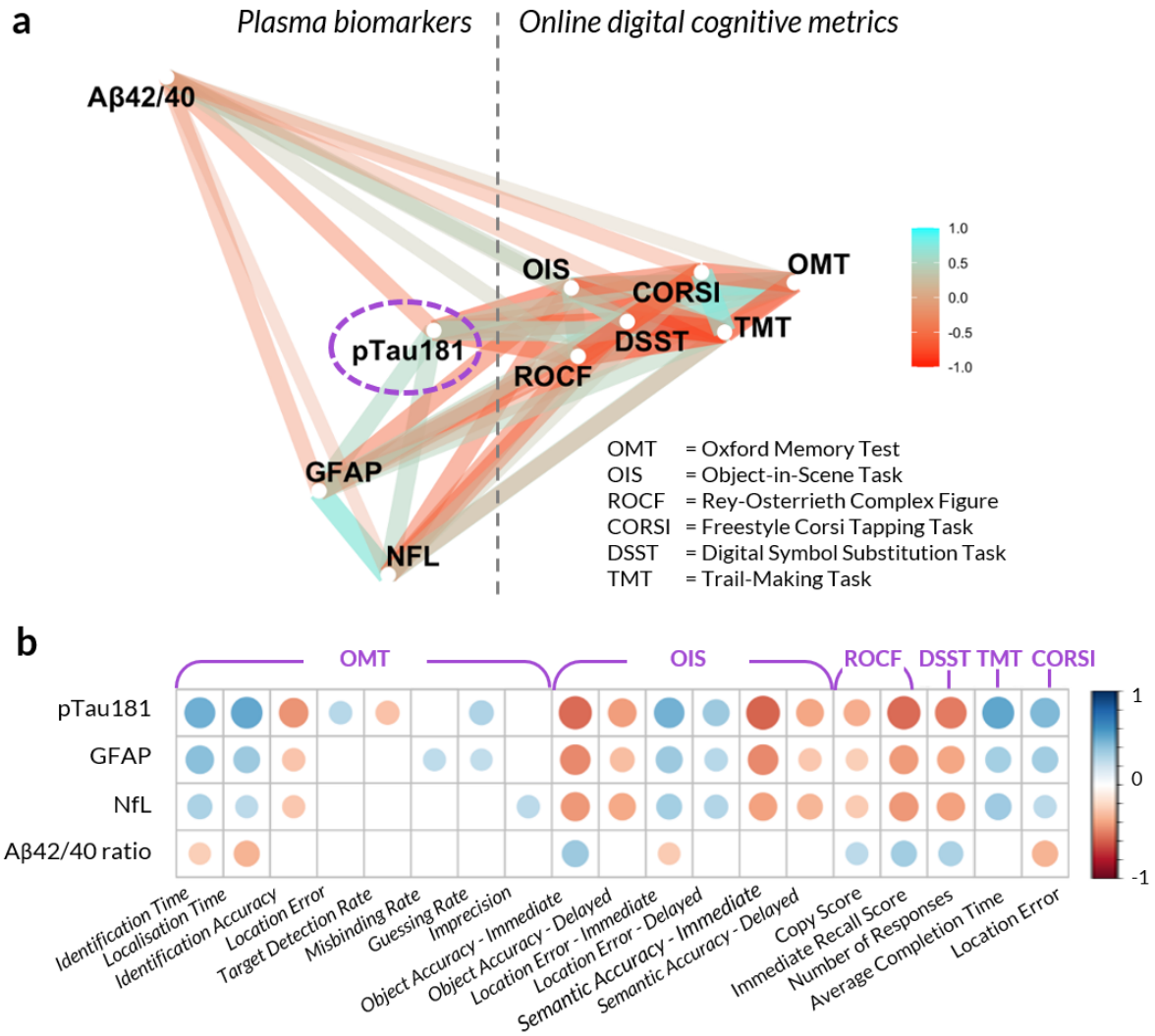

Figure 5. Which cognitive metric or plasma biomarker best predicts AD?

(a) Ranked biomarkers and digital cognitive metrics in predicting group AD or control. Plasma biomarkers are marked in orange, while online cognitive metrics are marked in black. All metrics were significant in group classification. OMT = Identification accuracy of the Oxford Memory Task, OIS = Object Identification Accuracy in Immediate Recall of the Object-in-Scene Memory Task, ROCF = recall of the Rey–Osterrieth Complex Figure, DSST = number of correct responses of the Digit Symbol Substitution Task, TMT = average reaction time of the Trail Making Test, CORSI = average location error of the Freestyle Corsi Block Task. **(b)** ROC curves for group classification. Light blue shows the combined model with pTau181, ROCF and OIS, with black line indicates ROCF alone model and orange shows the model pTau alone. AUC = area under the curve, CI = confidence interval.

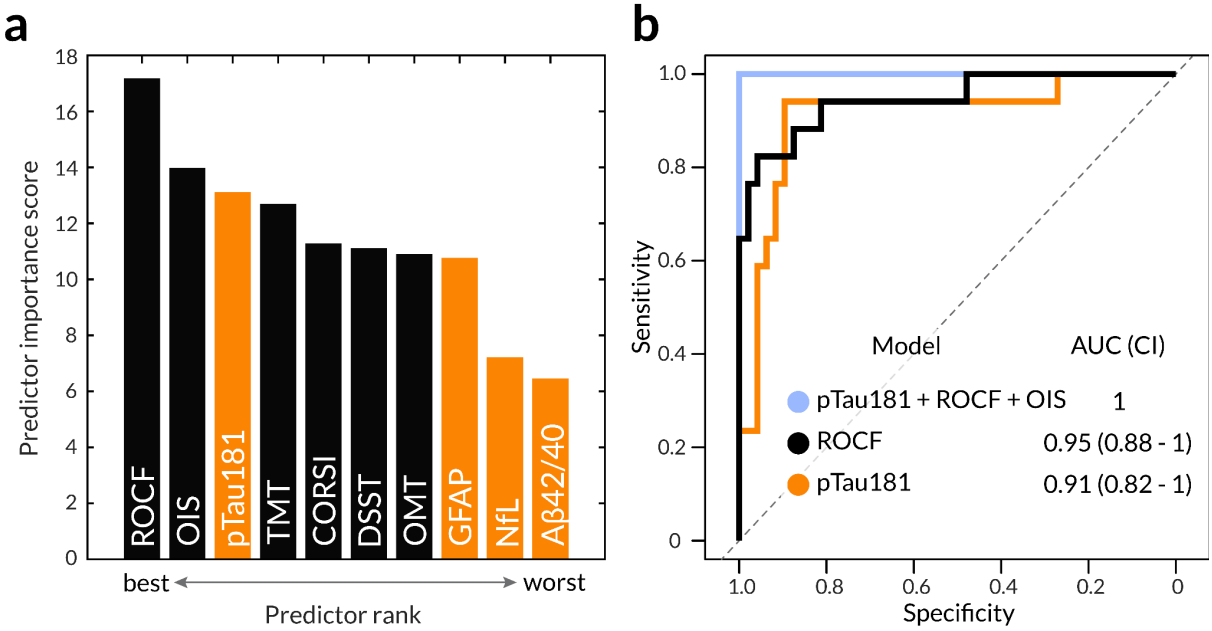

Table 1. Demographics, plasma biomarkers, standard cognitive metrics and questionnaire-derived apathy and depression scores.

All metrics are reported in group mean and standard deviation. * indicates p-values below <0.001, n.s. means not significant. The effect size for group comparison is rank-biserial correlation coefficient. ACE: The Addenbrooke's Cognitive Examination-III. AMI: Apathy Motivation Index. GDS: Geriatric Depression Scale.

Dimensions	Metrics	AD (n= 46)	EHC 1 (n= 53)	EHC 2 (n=256)	AD vs EHC 1	
					p value	Effect size
Demographics	Age	68.3 (10.2)	68.6 (7.0)	63.2 (8.2)	n.s.	0.021
	Gender (M/F)	23/23	24/29	126/130	n.s.	0.047
	Education	14.5 (3.5)	15.8 (3.1)	15.9 (3.6)	n.s.	0.199
Plasma biomarkers	pTau181 (pg/ml)	5.413 (3.4)	2.563 (1.3)	not applicable	* < 0.001	-0.70
	GFAP (pg/ml)	224.742 (117.9)	111.796 (57.9)		* < 0.001	-0.69
	NfL (pg/ml)	29.740 (19.1)	17.128 (11.0)		* < 0.001	-0.58
	Aβ42/40 ratio	0.059 (0.01)	0.067 (0.01)		* < 0.001	0.53
	Aβ42 (pg/ml)	6.640 (1.5)	7.154 (1.5)		0.06	0.22
	Aβ40 (pg/ml)	115.241 (25.7)	108.393 (19.9)		n.s.	-0.127
ACE	Total score	63.5 (20.0)	97.4 (2.0)	not applicable	* < 0.001	0.93
	Attention	12.0 (4.4)	17.2 (0.3)		* < 0.001	0.88
	Memory	11.7 (6.1)	25.0 (1.1)		* < 0.001	0.95
	Fluency	7.4 (3.6)	13.2 (0.9)		* < 0.001	0.89
	Language	21.0 (4.9)	25.6 (0.7)		* < 0.001	0.78
	Visuospatial	11.9 (4.2)	15.8 (0.3)		* < 0.001	0.79
Questionnaire	AMI	1.5 (0.3)	1.2 (0.4)	1.5 (0.4)	0.002	-0.40

s	GDS	6.9 (1.9)	4.9 (1.1)	5.6 (2.0)	* < 0.001	-0.64

Table 2. Digital cognitive metrics from OCTAL.

All metrics were normalised by age-matched normative data (i.e. participants who are EHC 2). All metrics are reported in group mean and standard deviation. * indicates p-values below <0.001. The effect size for group comparison is rank-biserial correlation coefficient. See supplementary table 1 for the extended table, including all cognitive metrics and the rank-biserial correlation coefficients for group comparison with EHC 2. OMT: Oxford Memory Test; OIS: Object-in-Scene task; ROCF: Rey-Osterrieth Complex Figure; CORSI: Freestyle Corsi tapping task; DSST: Digital Symbol Substitution Task; TMT: Trail-Making task.

Digital Cognitive Tasks	Metrics	AD (n= 46)	EHC 1 (n= 53)	EHC 2 (n=256)	AD vs EHC 1		AD vs EHC 2	EHC 1 vs EHC 2
					p value	Effect size	p value	p value
OMT	Identification Accuracy (z)	-1.9 (2.0)	0.3 (0.9)	-0.3 (1.2)	* < 0.001	0.66	* < 0.001	* < 0.001
	Location Error (z)	3.1 (2.6)	-0.1 (1.0)	0.4 (1.5)	* < 0.001	-0.77	* < 0.001	0.014
OIS	Object Identification Accuracy - Delayed Recall (z)	-2.4 (1.3)	0.0 (1.0)	-0.2 (1.0)	* < 0.001	0.86	* < 0.001	n.s.
	Location Error - Delayed Recall (z)	4.4 (3.5)	0.2 (1.0)	0.3 (1.4)	* < 0.001	-0.76	* < 0.001	n.s.
ROCF	Copy Raw Score (%)	73.6 (24.0)	90.6 (2.9)	89.6 (5.3)	* < 0.001	0.7	* < 0.001	n.s.
	Copy Score (z)	-6.0 (7.9)	0.1 (0.9)	-0.4 (1.9)	* < 0.001	0.69	* < 0.001	n.s.
	Immediate Recall Raw Score (%)	49.5 (20.2)	85.1 (7.8)	82.1 (9.5)	* < 0.001	0.92	* < 0.001	0.024
	Immediate Recall Score (z)	-5.2 (3.1)	0.1 (1.0)	-0.4 (1.3)	* < 0.001	0.92	* < 0.001	0.006
CORSI	Mean Location Error (z)	2.8 (2.3)	-0.0 (0.9)	0.4 (1.5)	* < 0.001	-0.8	* < 0.001	n.s.
DSST	Number of Correct	-2.5 (1.6)	-0.2 (0.9)	-0.1 (1.2)	* < 0.001	0.82	* < 0.001	n.s.

Digital Cognitive Tasks	Metrics	AD (n= 46)	EHC 1 (n= 53)	EHC 2 (n=256)	AD vs EHC 1		AD vs EHC 2	EHC 1 vs EHC 2
					p value	Effect size	p value	p value
	Responses (z)							
TMT	Average Completion Time (z)	10.8 (15.5)	0.3 (1.4)	0.2 (1.3)	* < 0.001	-0.89	* < 0.001	n.s.