

# Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease

David Berron<sup>1,2</sup>, Emil Olsson<sup>1</sup>, Felix Andersson<sup>3</sup>, Shorena Janelidze<sup>1</sup>, Pontus Tideman<sup>1,3</sup>,  
Emrah Düzel<sup>2,3,4</sup>, Sebastian Palmqvist<sup>1,5</sup>, Erik Stomrud<sup>1,5</sup>, Oskar Hansson<sup>1,5</sup>

<sup>1</sup> *Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, 223 62 Lund, Sweden*

<sup>2</sup> *German Center for Neurodegenerative Diseases, 39120 Magdeburg, Germany*

<sup>3</sup> *Institute for Cognitive Neurology and Dementia Research, Otto-von-Guericke University, Magdeburg, Germany*

<sup>4</sup> *Institute of Cognitive Neuroscience, University College London, United Kingdom*

<sup>5</sup> *Memory Clinic, Skåne University Hospital, 205 02 Malmö, Sweden*

## Corresponding authors:

David Berron

[david.berron@dzne.de](mailto:david.berron@dzne.de)

Clinical Cognitive Neuroscience Group

German Center for Neurodegenerative Diseases

Leipziger Straße 44

39120 Magdeburg, Germany

Oskar Hansson

[oskar.hansson@med.lu.se](mailto:oskar.hansson@med.lu.se)

Clinical Memory Research Unit

Department of Clinical Sciences Malmö

Lund University

Sölvegatan 19

223 62 Lund, Sweden

**Tables:** 3

**Figures:** 5

**NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## **Abstract**

**INTRODUCTION:** Remote unsupervised cognitive assessments have the potential to complement and facilitate cognitive assessment in clinical and research settings.

**METHODS:** Here we evaluate the usability, validity and reliability of unsupervised remote memory assessments via mobile devices in individuals without dementia from the Swedish BioFINDER-2 study and explore their prognostic utility regarding future cognitive decline.

**RESULTS:** Usability was rated positively; remote memory assessments showed good construct validity with traditional neuropsychological assessments and were significantly associated with tau-PET and downstream MRI measures. Memory performance at baseline was associated with future cognitive decline and prediction of future cognitive decline was further improved by combining remote digital memory assessments with plasma p-tau217. Finally, retest reliability was moderate for a single assessment and good for an aggregate of two sessions.

**DISCUSSION:** Our results demonstrate that unsupervised digital memory assessments might be used for diagnosis and prognosis in Alzheimer's disease, potentially in combination with plasma biomarkers.

**Keywords:** Alzheimer's disease, digital cognitive markers, memory, smartphone-based unsupervised assessments, blood-based biomarkers, plasma markers

## **Introduction**

While there has been significant progress in fluid and neuroimaging biomarkers to detect pathological changes in Alzheimer's disease (AD), cognitive measures in health care and clinical trials were initially designed to detect overt cognitive impairment and novel developments still lag behind (Hansson, 2021; Hansson et al., 2023). This is in stark contrast to recent discoveries on the functional architecture of episodic memory and its relationship to the spatial progression patterns of AD pathology (Berron et al., 2021; Maass et al., 2019). Recent work on the functional neuroanatomy of episodic memory have highlighted memory networks in the medial temporal lobe (MTL) and the neocortex that are involved in specific memory functions and are affected in different stages of AD (Berron et al., 2021; Grothe et al., 2017; Ranganath and Ritchey, 2012; Ritchey and Cooper, 2020). Episodic memory requires pattern separation and completion processes that are primarily mediated by MTL subregions. While the dentate gyrus is involved in pattern separation (Bakker et al., 2008; Berron et al., 2016) and reduces memory interference between similar events, the hippocampal Cornu Ammonis 3 (CA3) mediates pattern completion processes in interplay with neocortical regions (Grande et al., 2019). Within the hippocampal entorhinal circuitry there exist partly segregated pathways where object information is primarily provided from the perirhinal cortex via the anterior-lateral entorhinal cortex. Spatial information via the parahippocampal cortex is transferred additionally through the posterior parts of the entorhinal cortex (Berron et al., 2019, 2018; Grande et al., 2022; Maass et al., 2019, 2015; Schröder et al., 2015). Taken together, there is converging evidence that short-term mnemonic discrimination of object and scene representations, in addition to long-term memory, is impaired in the predementia stages of AD (Grande et al., 2021).

In addition, traditional neuropsychological assessment suffers from significant limitations such as high participant burden and impracticality of implementing test formats such as frequent test repetitions or long-term delay formats. Thus, traditional neuropsychological assessments become increasingly difficult for clinical trials in preclinical and early-stage symptomatic AD populations which gradually implement decentralized clinical trial structures for case-finding and monitoring, for example TRAILBLAZER-ALZ 3 (ClinicalTrials.gov Identifier: NCT05026866). In this context, unsupervised and remote digital cognitive assessments via smartphones and tablets offer a promising avenue to improve case-finding, monitoring and prognosis in both clinical trial and health-care settings (Nicosia et al., 2022b; Öhman et al., 2021, 2022; Papp et al., 2021). The aim of the present study was to evaluate the feasibility and usability of unsupervised and remote digital memory assessments, their construct validity in reference to traditional neuropsychological assessments, their retest reliability as well as their relationship with fluid and imaging biomarkers of AD. To that end, we implemented two non-verbal memory tasks based on recent insights into the functional anatomy of episodic memory,

available on the neotiv digital platform (<https://www.neotiv.com/en>) (Berron et al., 2022; Öhman et al., 2022), in a subset of the Swedish BioFINDER-2 study.

## **Methods**

### ***Recruitment into smartphone-based add-on study***

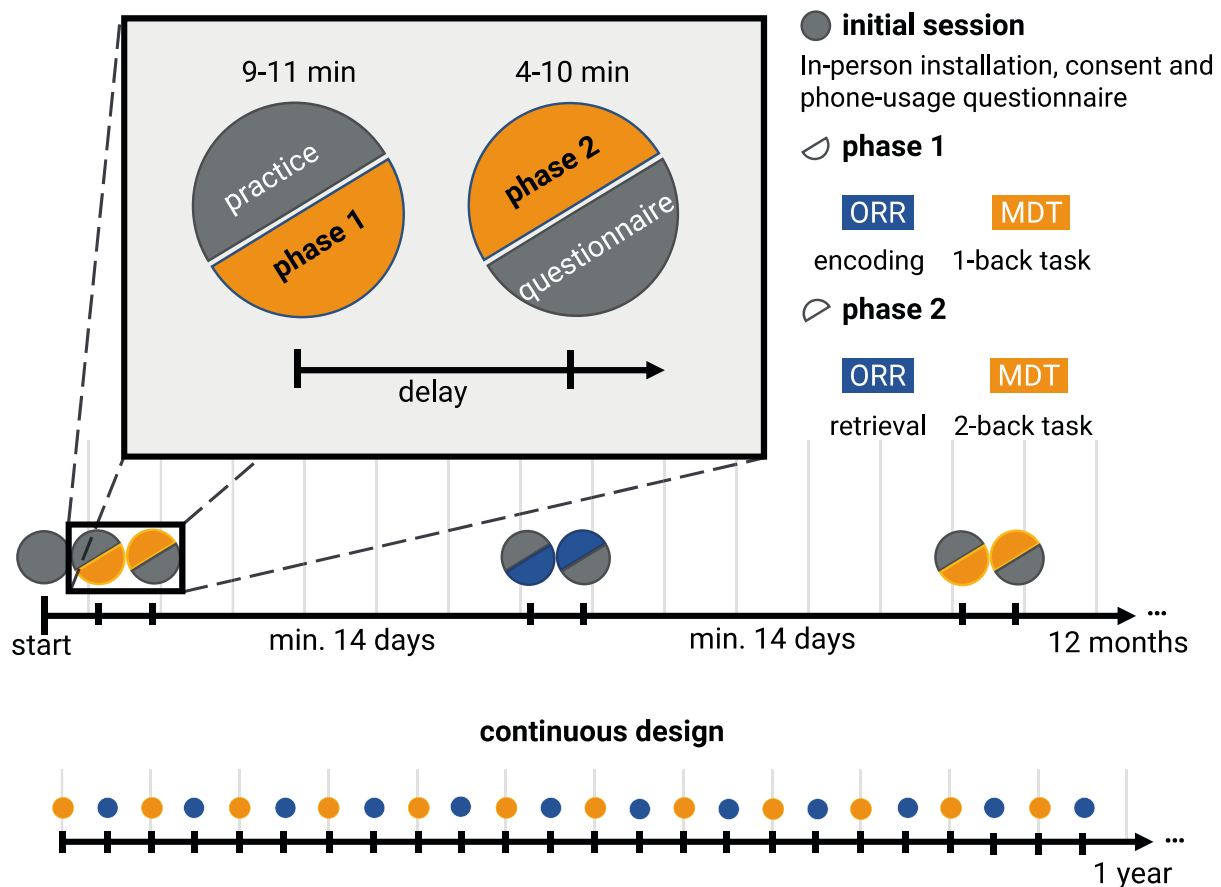
A total of 187 individuals who participated in the 7 Tesla MRI study arm of the Swedish BioFINDER2 study were offered participation in biweekly smartphone-based remote cognitive assessments across a 12-month period. 130 individuals (70%) gave written informed consent to participate.

A brief pen-and-paper questionnaire about experience with smartphones and mobile apps was completed and most participants downloaded and installed the neotiv app directly after the MRI session and completed the first phase of the first digital memory assessment session on site. A comprehensive user manual was distributed which included download and installation instructions and those individuals that did not download and install the app on-site successfully completed it without supervision from their home. The study was approved by the Regional Ethics Committee in Lund and the Swedish Ethical Review Authority.

Following completion of the initial app-hosted cognitive assessment, participants were notified to complete memory assessments every two weeks for 12 months. Each of the assessments consisted of a two-phase session. The two phases were either two halves of mnemonic discrimination, or encoding and retrieval phases of object-in-room recall (see details of the tasks below). Each phase took less than 10 minutes. Thus, overall participants could complete up to 12 sessions for each memory task within a period of approximately 12 months. In order to minimize potential practice effects from repeated testing, 12 independent difficulty-matched parallel test versions for each memory task were used. After each task completion, participants indicated their subjective task performance and their concentration level on a 5-point scale (1=very bad, 2=bad, 3=moderate, 4=good and 5=very good), and whether they had been distracted during the task (yes/no). Of note, only the first valid session of each participant was used in these cross-sectional analyses.

Push notifications were used to notify individuals about available tasks and to remind them daily for five consecutive days in case the tasks had not been initiated. At the beginning of each test session, participants were asked to go to a quiet environment, wear their glasses if needed and to adjust their screen's brightness to see the pictures clearly. They also received a short practice session for the initial test session as well as for all future sessions.

At the end of the study, 40 participants partook in telephone interviews to provide feedback on usability and user experience throughout the study. The full study timeline can be seen in Figure 1.



**Figure 1: Timeline of the study protocol.** Participants enlisted for a 12-month study of biweekly remote and unsupervised memory assessments of the Mnemonic Discrimination Task for Objects and Scenes (MDT-OS) and Objects-in-Room-Recall Task (ORR). In the initial in-clinic session, they gave consent, got a brief introduction, answered a short questionnaire on their phone usage, and completed the first task. Every second week, they received a short training session, followed by phase 1 of their respective task: encoding for ORR, and 1-back task for MDT-OS. After finishing phase 1, they were notified when the next phase was available, and could perform it straightaway or postpone if inconvenient; that is, there was a minimum delay of 24 h for the MDT-OS and a minimum delay of 60 minutes for the ORR, but it was often extended by the participants. Phase 2 consisted of retrieval for ORR, and 2-back task for MDT-OS. It was followed by ratings regarding concentration, distraction throughout and subjective difficulty of the task.

### **Smartphone-based memory assessments**

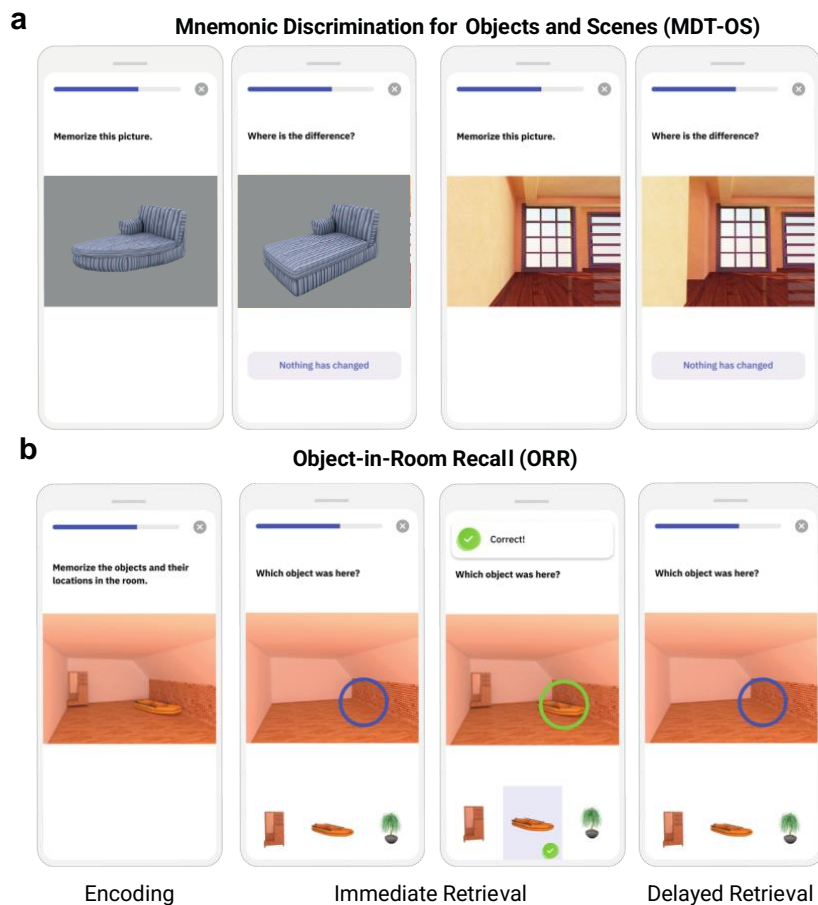
#### *Mnemonic Discrimination of Objects and Scenes (MDT-OS)*

In this continuous recognition task, individuals were presented with computer-generated images of various indoor objects and empty rooms, which were either exactly repeated, or slightly altered (see FigureA). Participants had to indicate whether an image was an identical repetition of a previous image (tap on a button), or had been modified (tap on the location of change). One session consisted of 64 image pairs (32 object pairs, 32 scene pairs), half of which were modified, and half of which were repeated. To optimize engagement while minimizing the subjective burden of participating, each session was split into two phases and automatically scheduled on two consecutive days with a 24-hour delay between phases. The first phase was presented as a one-back task while the second phase was presented as a two-back task. The MDT-OS has been designed to tax hippocampal pattern separation; a memory mechanism needed to discriminate between similar memories. Earlier studies using functional magnetic resonance imaging have shown that especially subregions in the

human medial temporal lobe are involved in this task (Berron et al., 2019, 2018; Maass et al., 2019). The test provides a hit rate, a false alarm rate and a corrected hit rate for both the object and scene condition. The averaged corrected hit rate across the scene and the object condition is used as the main outcome measure in the following analyses.

### *Object-In-Room Recall (ORR)*

In this task, participants had to memorize a spatial arrangement of two objects within a room. Following the encoding phase, a blue circle highlighted the previous position of one of the objects in the same but now empty room and the participant had to identify the correct object from a selection of three in an immediate retrieval phase (see Figure 2B). Among the three possible objects, one was the correct one that was previously shown in the now highlighted position (target), another belonged to the same room but had previously been shown at a different position (correct source distractor), and the third had previously been shown in a different room (incorrect source distractor). They learned 25 such object-scene associations. After 60 minutes, the participant was notified via push notification to complete an identical but delayed retrieval phase with a randomized stimulus order. The ORR has been designed to tax hippocampal pattern completion, a memory mechanism needed to restore full memories from partial cues (Berron et al., 2023; Grande et al., 2019). In the test, the assessment of recall is graded and allows to separate correct episodic recall from incorrect source memory. Thus, correct recall excludes the choice of an object that was present in the same room but at a different location (wrong source memory for specific location), and an object that was not present in the room but nevertheless associated with another room during encoding (wrong source memory for overall location). The test provides an immediate recall (ORR-IR; 0-25) and a delayed recall score (ORR-DR; 0-25). Here we use the ORR-DR.



**Figure 2: Smartphone-based memory tasks.** (A) Mnemonic Discrimination Test for Objects and Scenes (MDT-OS) and (B) Objects-in-Room-Recall (ORR) test.

### *Quality Control Procedures*

For this manuscript, we analyzed data up until the data release in July 2023. Recruitment of participants from the Swedish BioFINDER-2 study was performed between February 2019 and February 2022. Both memory tasks were presented in two phases, and we only included participants that had completed at least one full session, i.e., both phases. The following outcome measures were calculated from the individual trials of each memory task: For the MDT-OS, we calculated the corrected hit rate to adjust for response bias by subtracting the false alarm rate (percentage of modified images that were identified as old) from the hit rate (the percentage of repeated images that were identified as repeated). The resulting cognitive outcomes are thus the overall corrected hit rate for mnemonic discrimination (average across all object and scene trials) as well as corrected hit rates for the object and scene condition separately. For the ORR, we calculated the delayed retrieval accuracy. Regarding the sessions from the first wave, meaning each very first assessment of MDT-OS and ORR-DR, 6% of test sessions exceeded the threshold for missing responses (maximum of 25% of missing responses per session) and 17% of test sessions exceeded the maximum length of the delay period (>240 minutes) before filtering. These sessions were excluded during quality assessment. Excluded test sessions were

replaced by valid subsequent sessions where possible. As a result, 67% of test sessions we report here were from the first wave and 33% from subsequent test sessions.

### ***CSF and plasma analysis***

Cerebrospinal fluid (CSF)  $\beta$ -amyloid ( $A\beta$ )<sub>42</sub> and  $A\beta$ <sub>40</sub> concentrations were measured using the Elecsys assays (Roche Diagnostics). Participants were stratified for  $A\beta$  status using the CSF  $A\beta$ <sub>42/40</sub> ratio with a predefined cut-off of 0.08 (Quadalti et al., 2023). Plasma levels of p-tau<sub>217</sub> were quantified with the MSD-based immunoassay developed by Lilly Research Laboratories as previously described (Palmqvist et al., 2020).

### ***Imaging acquisition***

#### ***MRI***

T1-weighted images were acquired on a Siemens Prisma scanner (Siemens Medical Solutions, Erlangen, Germany) with a 64-channel head coil using an MPRAGE sequence (in-plane resolution=1×1mm<sup>2</sup>, slice thickness=1mm, TR=1900ms, TE=2.54ms, flip-angle=9°). Spontaneous BOLD oscillations were acquired with a gradient-echo planar sequence (eyes closed, in-plane resolution=3×3mm<sup>2</sup>, slice thickness=3.6mm, TR=1020ms, TE=30ms, flip-angle=63°, 462 dynamic scans, 7.85min).

#### ***Tau and $A\beta$ -PET***

All study participants underwent positron emission tomography (PET) scans on a digital GE Discovery MI scanner (General Electric Medical Systems). Approval for PET imaging was obtained from the Swedish Medical Products Agency. For tau-PET imaging, the participants were injected with 365±20 MBq of [<sup>18</sup>F]RO948, and LIST mode emission data was acquired for each scan 70–90 min ([<sup>18</sup>F]RO948) post injection.  $A\beta$ -PET imaging was performed on the same platform 90–110 min after the injection of ~185 MBq [<sup>18</sup>F]Flutemetamol (Cho et al., 2020).

### ***Imaging analysis***

#### ***ROI segmentation and estimates of volume and thickness***

Individual volume of the anterior and posterior hippocampus and median thickness of area 35 were defined on T1-weighted images (1x1x1mm<sup>3</sup> resolution) using ASHS-T1 (Xie *et al.*, 2019) and a multi-template thickness analysis pipeline (Xie et al., 2017). All subregional masks were visually assessed.



### ***SUVr measures***

*[<sup>18</sup>F]RO948*: Standardized uptake value ratio (SUVr) images were calculated for area 35 using an inferior cerebellar reference region (Baker et al., 2017). Partial volume correction (PVC) was performed using the geometric transfer matrix method (Rousset et al., 1998) extended to voxel-level using region-based voxel-wise correction (Thomas et al., 2011). *[<sup>18</sup>F]Flutemetamol*: A cortical composite SUVr as well as a measure of early amyloid deposition in the precuneus was calculated using the whole cerebellum as reference region (Thurfjell et al., 2014).

### ***Traditional cognitive measures***

As a measure representing episodic memory, we used the delayed 10-word list recall test from the Alzheimer's Disease Assessment Scale—Cognitive Subscale (ADAS-Cog; Rosen et al., 1984). The learning trial of the 10 words was repeated three times. After a distraction task (Boston Naming—15 items short version), the participant was asked to freely recall the 10 words (“delayed recall”). Delayed recall was scored as number of errors (*i.e.*, 10 minus correct recalled words), so that a higher score entailed worse memory performance. For global cognition, we used the Mini-Mental State Examination (MMSE; scale from 0–30). For attentional and executive function, we used the Symbol Digit Modalities Test (SDMT, one point for every correct answer within the response time of 90 seconds) and verbal fluency (Animal fluency; number of correct animals within the response time of 60 seconds). In addition, we used the corrected hit rates for object and scene memory derived from an on-site version of the test that was completed while lying in the fMRI scanner. In the end, we calculated a composite score similar to the Preclinical Alzheimer Cognitive Composite 5 (PACC5), the modified PACC (mPACC), using the average of the z-standardized scores of ADAS delayed recall (counted twice), SDMT, MMSE and verbal fluency.

### ***Statistical analysis***

Multiple regression analyses were carried out between PET measures, MTL subregional atrophy and unsupervised remote memory tests in R (version 4.1.2; [www.r-project.org](http://www.r-project.org); R Core Team, 2022). All models were adjusted for age, sex and intracranial volume (for models including volumes). Results were corrected for multiple comparisons using False-Discovery Rate correction ( $p < 0.05$ ) where appropriate. Robust regression models were estimated using iteratively re-weighted least squares (ILRS) using the MASS package (`rlm` function).

We extracted participant-specific mPACC slopes from linear mixed-effects models with random intercepts and slopes (using the `lme4` package for R) and mPACC score as the outcome and time (visit number) as the predictor. These participant-specific slopes were used as outcomes in a second set of linear regression models with plasma p-tau217 and remote memory measures as

predictors while adjusting for age, sex and years of education. For comparison, we also fit basic models using only the covariates, without cognitive markers or biomarkers. Models were evaluated using Akaike information criterion (AIC). A difference of 2 AICs is considered significant (Burnham and Anderson, 2004). Retest reliability was assessed for two individual sessions (Session 1 vs 2; on average after 7 weeks) and for two averaged sessions (mean of Session 1 and 2 vs. mean of Session 3 and 4; on average after 13 weeks). Intraclass correlation coefficients (ICC) and their 95% confident intervals were calculated based on single rater, absolute-agreement, two-way random-effects model. We consider ICC values less than 0.5 indicative of poor reliability, values between 0.5 and 0.75 indicative for moderate reliability, values between 0.75 and 0.9 indicative for good reliability, and values greater than 0.90 indicative for excellent reliability (Koo and Li, 2016).

### ***Data availability***

Anonymized data will be shared by request from a qualified academic investigator for the sole purpose of replicating procedures and results presented in the article and as long as data transfer is in agreement with EU legislation on the general data protection regulation and decisions by the Ethical Review Board of Sweden and Region Skåne, which should be regulated in a material transfer agreement.

## Results

### Participant sample

Here we considered the 130 study participants who opted to participate in the add-on smartphone-based study. While 20 did not complete any test session, another 11 were excluded during quality assurance as described above. As a result, 99 individuals who contributed at least one valid test session of the MDT-OS (see table 1) and 66 participants who contributed at least one valid test session of the MDT-OS and the ORR task (see supplementary table 2) could be included. Following initial filtering, we thus analyzed data from 99 participants in the MDT-OS task (48 CU A $\beta$ –, 28 CU A $\beta$ +, 9 MCI A $\beta$ – and 14 MCI A $\beta$ +) and from 66 participants in the ORR task (31 CU A $\beta$ –, 19 CU A $\beta$ +, 7 MCI A $\beta$ – and 9 MCI A $\beta$ +).

**Table 1: Participant demographics of the entire sample**

	CU A $\beta$ – (N=48)	CU A $\beta$ + (N=28)	MCI A $\beta$ – (N=9)	MCI A $\beta$ + (N=14)	Total (N=99)
Age (years)	60.8 (10.9)	69.8 (8.4)	64 (11.9)	67.8 (4.3)	64.6 (10.4)
Education (years)	13.4 (3.3)	12.9 (3.3)	12.7 (2.7)	12.5 (3.8)	13.1 (3.3)
Sex (% female)	54.2 %	57.1 %	55.6 %	35.7 %	52.5 %
MMSE	29 (1.2)	28.6 (1.6)	27 (2.1)	27.7 (1.5)	28.5 (1.6)

Displayed are mean values (standard deviations) unless otherwise stated

Abbreviations: N, number of participants; CU, Cognitively Unimpaired; MCI, mild cognitive impairment; MMSE, mini-mental state examination.

### Older participants show high acceptability of smartphone-based assessments

First of all, we were interested in the acceptability of the remote smartphone-based assessments. 127 participants completed a questionnaire about their phone usage. Out of those, 90% owned a smartphone, 62% said they would download apps by themselves, and 22% with assistance. Participants reported that they had been using a smartphone for 8.6 years on average (SD: 4.3 years, range: 1 – 25 years), and spend 2.3 hours per day using it (SD: 1.8 hours, range: 1–10 hours).

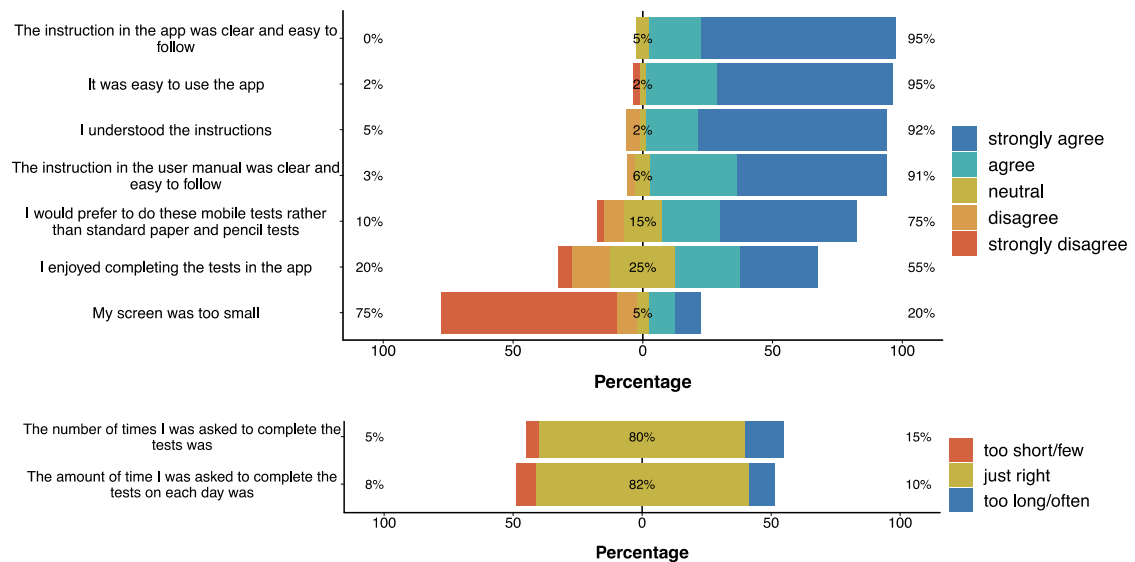
As reported above, 130 individuals were recruited into the add-on smartphone study (70% of those that were asked). The most frequent reason not to participate in the smartphone-based study was lack of time (too many assessments; 13%), followed by unfamiliarity with smartphone use and insecurity about their capabilities (7%), as well as not owning a smartphone with internet connection (7%). 73% did not further specify any reasons.

Across both cognitive tests, participants reported high average concentration levels (MDT-OS = 3.9; ORR = 3.9; scale 1–5, which translates to good concentration) during the task as well as moderately high subjectively rated task difficulty (MDT = 3.1; ORR = 3.4; scale 1–5, which translates to moderate subjectively rated performance). Both estimates were very similar across diagnostic groups (group average range in concentration from 3.7–4.1 and in subjective task difficulty from 2.9–3.5).

However, as expected, the second phase of the task (i.e., 2-back or delayed retrieval compared to one-back or immediate retrieval) was perceived as considerably more difficult (phase 1: 3.7; phase 2: 2.7). Across both tasks, 88% of participants reported no distractions during the task and distraction rates were higher in cognitively unimpaired individuals (on average 14% across both tasks) compared to MCI patients (on average 6% across both tasks).

Following filtering, the time between encoding and retrieval in the ORR test was on average 96 minutes (SD = 44). Mobile devices had a screen diagonal between 10.2–24.6 cm (mean 13.2 cm, SD = 3.4) indicating the use of smartphones as well as tablet computers.

Finally, we approached 40 participants for a telephone-based interview to learn about their experience with the study (see Figure 3). Over 90% of them found the tasks and instructions easy to understand, and the app easy to use. 75% said they would prefer the mobile tests over in-person paper-and-pencil tests. 80% did not experience their device’s screen as too small to see all the details in the tasks. 80% found the number of tests and their duration just right, and 82.5% rated their experience using the app 7 or higher on a 10-point scale. There were only 10% of participants who had dropped out of the study earlier but still completed the phone interview. Those rated their experience with the app lower than participants who remained in the study, and they perceived the number and duration of assessments as too long.



**Figure 3: Acceptability and User experience.** Results from telephone-based interviews with 40 participants focusing on their overall experience with the remote and unsupervised study.

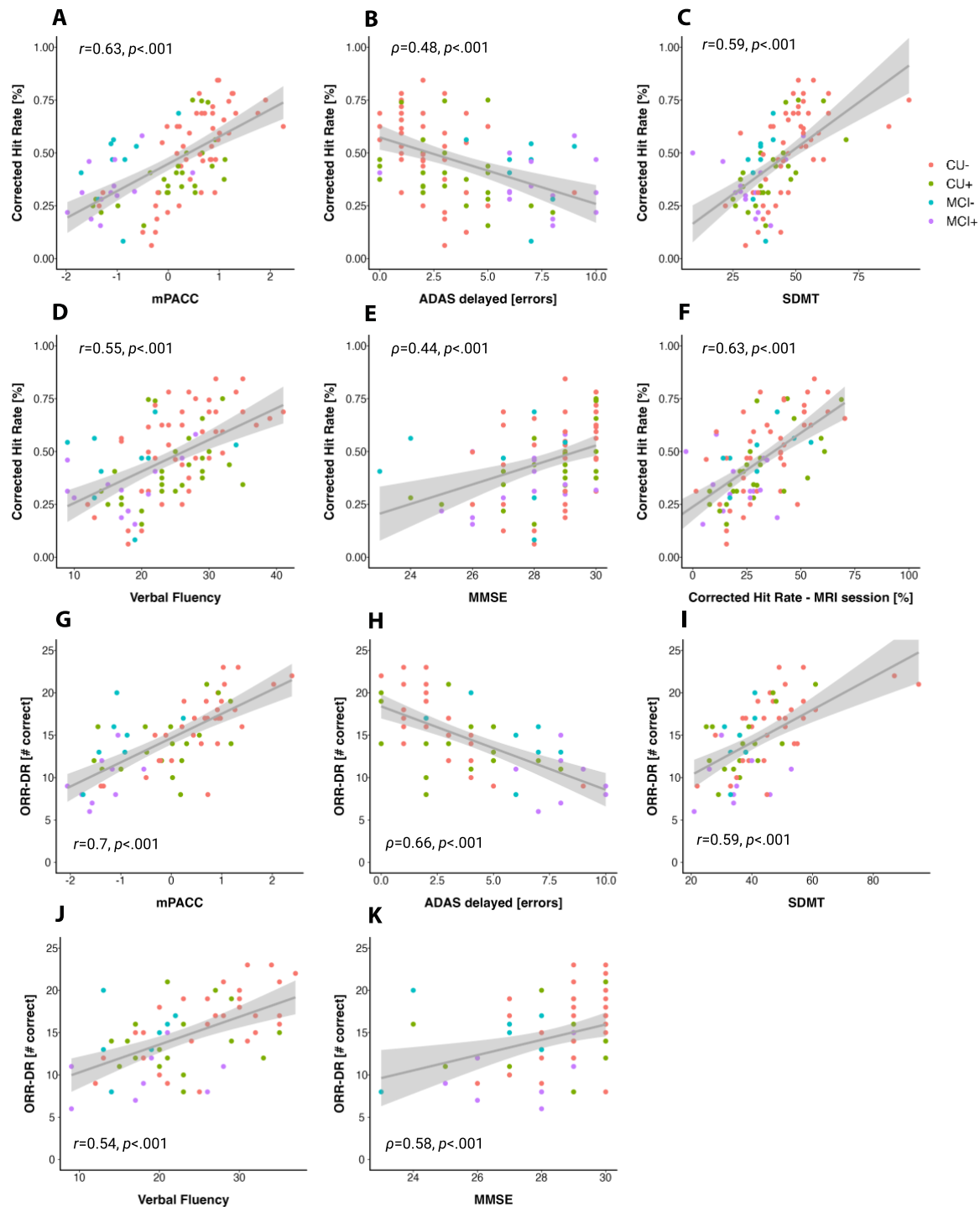
### ***Unsupervised memory assessments via mobile devices reflect in-clinic cognitive scores***

Next, we were interested whether the performance in the unsupervised and remote memory assessments was associated with traditional in-clinic cognitive measures. In order to understand the relationship with demographics and potential confounds, we first calculated multiple regression

models with age, sex, and years of education to identify the associations with both memory paradigms. In addition, we included time-to-retrieval, i.e., the time period between memory encoding and retrieval, as a covariate for the ORR. For MDT-OS, age and sex were significant predictors for the object, and only age was a significant predictor for the scene condition where higher age and female sex were associated with worse task performance (MDT-O:  $\beta_{\text{age}} = -0.009, p < .001$ ;  $\beta_{\text{sex}} = -0.08, p = 0.03$ ;  $\beta_{\text{education}} = 0.005, p = 0.34$ ; MDT-S:  $\beta_{\text{age}} = -0.008, p < .001$ ;  $\beta_{\text{sex}} = -0.06, p = 0.13$ ;  $\beta_{\text{education}} = 0.008, p = 0.186$ ). For the ORR, again only age was associated with task performance ( $\beta_{\text{age}} = -0.23, p < .001$ ;  $\beta_{\text{education}} = 0.09, p = 0.555$ ;  $\beta_{\text{sex}} = 0.39, p = 0.695$ ;  $\beta_{\text{delay}} = -0.09, p = 0.890$ ).

Next, we assessed the construct validity of the outcomes of the unsupervised memory assessments by analyzing the relationship with several in-clinic and supervised cognitive assessments (see Figure 4). This included an in-scanner version of the MDT-OS which was completed by the same participants during an fMRI scan. While this was a similar task (mnemonic discrimination of objects and scenes; (Berron et al., 2019, 2018; Maass et al., 2019), task administration (controller button presses instead of a touch screen) and environment (lying in an MRI scanner vs. completing tasks at home) differed compared to the remote unsupervised setting. Pearson correlation coefficients revealed a strong relationship between the MDT-OS corrected hit rate across both settings ( $r = 0.63, p < .001$ ; MDT-O:  $r = 0.43, p < .001$ ; MDT-S:  $r = 0.49, p < .001$ ). Next, we were interested whether MDT-OS and ORR-DR were associated with an in-clinic measure of memory (ADAS delayed recall) and a composite score that has been shown to be sensitive to early AD (mPACC) as well as the Symbol Digit Modalities Test (SDMT), verbal fluency and the Mini-Mental State Examination (MMSE). Due to ceiling effects of ADAS delayed recall and the MMSE in early Alzheimer's disease, we assessed Spearman rank correlation rho. All measures were significantly associated with the MDT-OS corrected hit rates (*ADASdelayed*:  $\rho = 0.48, p < .001$ ; MDT-O:  $\rho = 0.53, p < .001$ ; MDT-S:  $\rho = 0.33, p < .001$ ; *MMSE*:  $\rho = 0.44, p < .001$ ; MDT-O:  $\rho = 0.35, p < .001$ ; MDT-S:  $r = 0.41, p < .001$ ; *VerbalFluency*:  $r = 0.55, p < .001$ ; MDT-O:  $r = 0.49, p < .001$ ; MDT-S:  $r = 0.49, p < .001$ ; *SDMT*:  $r = 0.59, p < .001$ ; MDT-O:  $r = 0.55, p < .001$ ; MDT-S:  $r = 0.50, p < .001$ ). Similarly, ORR delayed retrieval performance was most strongly associated with ADAS delayed recall but also with all other outcomes (*ADASdelayed*:  $\rho = 0.66, p < .001$ ; *MMSE*:  $\rho = 0.41, p < .001$ ; *VerbalFluency*:  $r = 0.54, p < .001$ ; *SDMT*:  $r = 0.58, p < .001$ ).

The mPACC score was associated with MDT-OS (mPACC:  $r = 0.63, p < .001$ ; MDT-O:  $r = 0.61, p < .001$ ; MDT-S:  $r = 0.51, p < .001$ ) and ORR delayed retrieval performance ( $r = 0.7, p < .001$ ). All relationships above survived corrections for multiple comparisons using FDR correction.



**Figure 4: Construct validity of the Mnemonic Discrimination Task for Objects and Scenes (MDT-OS) and the Object-in-Room Recall (ORR).** Scatter plots showing relationships of the MDT-OS Corrected Hit Rate with the (A) modified Preclinical Alzheimer’s Cognitive Composite, (B) errors in the delayed 10-word list recall test from ADAS-cog, (C) Symbol Digit Modalities Test, (D) Animal Fluency, (E) Mini Mental State Examination as well as the (F) Mnemonic Discrimination Task for Objects and Scenes Corrected Hit Rate from an on-site based task version that was performed during an MRI scan. Furthermore, scatter plots show relationships of the ORR Delayed Recall Score (ORR-DR) with the (G) modified Preclinical Alzheimer’s Cognitive Composite, (H) errors in the delayed 10-word list recall test from ADAS-cog, (I) Symbol Digit Modalities Test, (J) Animal Fluency and the (K) Mini Mental State Examination.

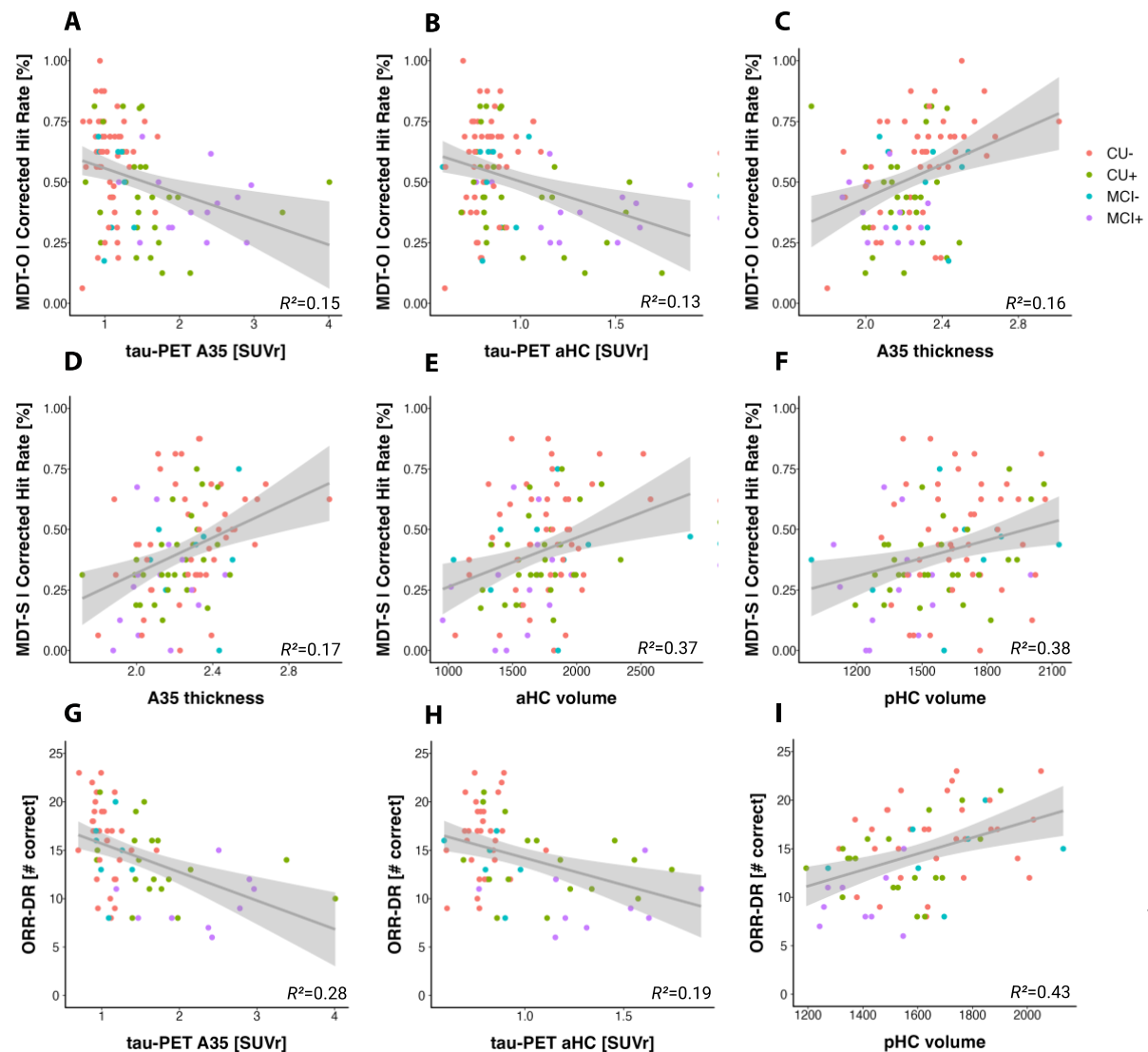
### ***Digital memory scores are associated with measures of AD pathology and medial temporal lobe atrophy***

Next, we were interested whether memory scores from remote and unsupervised assessments were associated with measures of AD pathology, namely A $\beta$  and tau pathology, as well as measures of atrophy (see Figure 5). Both memory tasks have been designed to tax medial temporal lobe-dependent memory processes (Berron et al., 2019, 2018; Maass et al., 2019). Thus, we were interested in the relationship with measures of tau pathology representing early disease stages. Following recent findings on the relationship of memory performance with measures of MTL tau pathology and atrophy (Berron et al., 2021; Flores et al., 2022; Sanchez et al., 2021), we selected tau-PET SUVR in area 35 and the anterior and posterior hippocampus as well as median thickness of area 35 and the volume of the anterior and posterior hippocampus. Regarding A $\beta$ , we selected on SUVR in the precuneus representing early  $\beta$ -amyloid deposition (Palmqvist et al., 2017).

Regarding tau-PET, MDT-O, but not MDT-S, was significantly associated with A35 tau-PET SUVR (MDT-O:  $\beta=-0.71$ , SE=0.32,  $p=0.029$ ; MDT-S:  $\beta=-0.32$ , SE=0.29,  $p=.269$ ) and anterior hippocampal tau-PET SUVR when accounting for participant age and sex (MDT-O:  $\beta=-0.44$ , SE=0.14,  $p=.003$ ; MDT-S:  $\beta=-0.017$ , SE=0.13,  $p=.90$ ). Neither measure was associated with posterior hippocampal tau-PET SUVR (MDT-O:  $\beta=-0.25$ , SE=0.13,  $p=.055$ ; MDT-S:  $\beta=-0.001$ , SE=0.12,  $p=.966$ ). Likewise, ORR-DR performance was significantly associated with A35 tau-PET SUVR ( $\beta=-0.06$ , SE=0.02,  $p=.007$ ) and anterior ( $\beta=-0.02$ , SE=0.009,  $p=.016$ ) but not posterior hippocampal tau-PET SUVR ( $\beta=-0.01$ , SE=0.008,  $p=.173$ ) when accounting for participant age, sex and time between encoding and retrieval. These relationships survived corrections for multiple comparisons using FDR correction. Similarly, regarding A $\beta$ -PET, MDT-O and ORR-DR performance, but not MDT-S, were significantly associated with A $\beta$ -PET SUVR in precuneus (MDT-O:  $\beta=-0.29$ , SE=0.13,  $p=.03$ ; MDT-S:  $\beta=-0.12$ , SE=0.12,  $p=.318$ ; ORR:  $\beta=-0.02$ , SE=0.008,  $p=.014$ ). However, A $\beta$ -PET SUVR relationships did not hold when accounting for A35 tau-PET SUVR.

With respect to MTL measures of atrophy, MDT-O and MDT-S were significantly associated with area 35 thickness (MDT-O:  $\beta=0.23$ , SE=0.11,  $p=.043$  [robust regression:  $F=5.34$ ,  $p=.022$ ]; MDT-S:  $\beta=0.22$ , SE=0.1,  $p=.025$ ) but only MDT-S was associated with anterior (MDT-O:  $\beta=282.4$ , SE=155.4,  $p=.073$ ; MDT-S:  $\beta=401.9$ , SE=135.4,  $p=.004$ ) and posterior hippocampal volume (MDT-O:  $\beta=215.53$ , SE=111.6,  $p=.057$ ; MDT-S:  $\beta=202.7$ , SE=99.8,  $p=.045$ ) when accounting for age, sex and intracranial volume. ORR-DR performance was significantly associated with posterior hippocampal volume ( $\beta=14.8$ , SE=5.7,  $p=.013$ ) but neither anterior hippocampal volume ( $\beta=14.7$ , SE=7.8,  $p=.064$ ) nor area 35 thickness ( $\beta=0.008$ , SE=0.006,  $p=.175$ ) when accounting for age, sex, intracranial volume and time between encoding and retrieval. The relationships of the MDT-O and MTL atrophy measures did not survive corrections for multiple comparisons using FDR correction.

Thus, both memory tasks showed sensitivity to measures of AD pathology as well as MTL atrophy as a measure of neurodegeneration.



**Figure 5: Criterion validity of the Mnemonic Discrimination Task for Objects and Scenes (MDT-OS) and the Object-in-Room Recall – Delayed Recall Score (ORR-DR).** Scatter plots showing relationships between the MDT-O and tau-PET uptake in (A) area 35 and (B) the anterior hippocampus, as well as (C) cortical thickness in area 35. Panel (D), (E), and (F) show relationships of the MDT-S with area 35 cortical thickness, anterior and posterior hippocampal volume. Finally, panel (G), (H) and (I) show illustrate relationships of the ORR-DR and tau-PET uptake in area 35 and the anterior hippocampus as well as the volume of the posterior hippocampus. Abbreviations: A35, area 35; aHC, anterior hippocampus; pHC, posterior hippocampus; MDT-OS; Mnemonic Discrimination task for Objects and Scenes; ORR-DR, Object-In-Room Recall – Delayed Recall.

### **Combination of blood-based biomarker and digital cognitive marker predicts future cognitive decline**

Next, we examined whether baseline scores from unsupervised memory assessments were associated with future cognitive decline rates in the mPACC and whether a combination of a blood-based biomarker and digital cognitive markers would outperform individual measures in predicting future cognitive decline. Recent work has shown that plasma p-tau217 was associated with cognitive decline as well as disease progression (Mattsson-Carlgrén et al., 2023; Palmqvist et al., 2021). To that end, we derived participant-specific mPACC slopes from linear mixed-effects model with random



intercepts and slopes and ran hierarchical linear regression models with plasma p-tau217 and remote memory measures as predictors in order to test which model best predicted future mPACC decline across 5 years.

Adjusting for covariates, cognitive outcomes for MDT-S and ORR-DR as well as p-tau217 but not MDT-O were significantly associated with mPACC slopes (see Table 2 and 3, see also supplementary table 1 for individual longitudinal linear mixed-effects models for each marker).

We next aimed to define an optimal combination to predict mPACC slopes. The best combination model for prediction of mPACC slopes (i.e., the model with the lowest AIC) included plasma p-tau217 ( $\beta$  [SE] = -0.026 [0.004];  $P < .001$ ), MDT-S ( $\beta$  [SE] = 0.157 [0.023];  $P < .001$ ), age (with higher age associated with worse mPACC slopes;  $\beta$  [SE] = -0.002 [0.001];  $P = < .001$ ), sex (with male sex associated with worse mPACC slopes;  $\beta$  [SE] = 0.039 [0.009];  $P < .001$ ), and education ( $\beta$  [SE] = -0.001 [0.001];  $P = .307$ ; AIC for the overall model: -941;  $R^2 = 0.29$ , see Table 2 “Digital Cognitive Marker + BBM”). This model was also better than both the Digital Cognitive Marker and the BBM model as indicated by a lower AIC.

Similarly, in a smaller sample of individuals that completed both the MDT-OS and the ORR-DR (see supplementary table 2 for sample characteristics), the best combination model for prediction of mPACC slopes included plasma p-tau217 ( $\beta$  [SE] = -0.03 [0.004];  $P < .001$ ), ORR-DR ( $\beta$  [SE] = 0.012 [0.002];  $P < .001$ ), MDT-S ( $\beta$  [SE] = 0.142 [0.029];  $P < .001$ ), age ( $\beta$  [SE] = 0.001 [0.001];  $P = .252$ ), sex (with male sex associated with worse mPACC slopes;  $\beta$  [SE] = 0.031 [0.012];  $P < .001$ ), and education (with higher education associated with worse mPACC slopes; ( $\beta$  [SE] = -0.006 [0.002];  $P < .001$ ; AIC for the overall model: -606;  $R^2 = 0.38$ , see Table 3 “Digital Cognitive Marker + BBM”). This model was also better than both the Digital Cognitive Marker and the BBM model with respect to lower AIC.

**Table 2: Results of multiple regression models in the entire sample (n=85) predicting future cognitive decline including baseline plasma and digital remote markers as predictors, demographic covariates, and mPACC slopes as an outcome.**

Predictor	Null model			Digital Cognitive Marker			BBM			Digital Cognitive Marker + BBM		
	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p
(intercept)	0.250	0.170 – 0.330	<0.001	0.092	-0.003 – 0.188	0.058	0.244	0.168 – 0.319	<0.001	0.126	0.045 – 0.206	0.002
Age [years]	-0.004	-0.005 – -0.003	<0.001	-0.002	-0.003 – -0.001	<0.001	-0.003	-0.004 – -0.002	<0.001	-0.002	-0.003 – -0.001	<0.001
Sex [female]	0.022	0.003 – 0.041	0.020	0.041	0.022 – 0.060	<0.001	0.024	0.006 – 0.042	0.008	0.039	0.021 – 0.057	<0.001
Education [years]	0.002	-0.001 – 0.005	0.256	-0.001	-0.004 – 0.002	0.574	0.001	-0.002 – 0.004	0.566	-0.001	-0.004 – 0.001	0.307
MDT-O				0.040	-0.019 – 0.099	0.179						
MDT-S				0.152	0.098 – 0.206	<0.001				0.157	0.111 – 0.203	<0.001
p-tau217							-0.028	-0.035 – -0.021	<0.001	-0.026	-0.034 – -0.019	<0.001
Observations	537			537			537			537		
R <sup>2</sup> /R <sup>2</sup> adj	0.144 / 0.139			0.217 / 0.210			0.226 / 0.221			0.287 / 0.281		
AIC	-846.013			-890.164			-898.508			-940.504		

**Table 3: Results of multiple regression models in a restricted sample (n=66) where both digital markers were completed predicting future cognitive decline including baseline plasma and digital remote markers as predictors, demographic covariates, and mPACC slopes as an outcome.**

Predictor	Null model			Digital Cognitive Marker			BBM			Digital Cognitive Marker + BBM		
	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p
(intercept)	0.287	0.182 – 0.392	<0.001	-0.161	-0.292 – -0.030	0.016	0.266	0.165 – 0.366	<0.001	-0.144	-0.265 – -0.023	0.020
Age [years]	-0.004	-0.005 – -0.003	<0.001	-0.000	-0.001 – 0.001	0.878	-0.003	-0.004 – -0.002	<0.001	0.001	-0.001 – 0.002	0.253
Sex [female]	0.018	-0.008 – 0.044	0.177	0.037	0.013 – 0.062	0.003	0.015	-0.009 – 0.040	0.220	0.031	0.008 – 0.053	0.008
Education [years]	-0.000	-0.004 – 0.003	0.813	-0.006	-0.010 – -0.002	0.001	-0.001	-0.005 – 0.003	0.586	-0.006	-0.010 – -0.003	<0.001
MDT-O				0.078	0.002 – 0.153	0.044						
MDT-S				0.098	0.031 – 0.166	0.004				0.142	0.085 – 0.200	<0.001
ORR-DR				0.012	0.008 – 0.016	<0.001				0.012	0.008 – 0.015	<0.001
p-tau217							-0.031	-0.041 – -0.021	<0.001	-0.030	-0.039 – -0.021	<0.001
Observations	347			347			347			347		
R <sup>2</sup> /R <sup>2</sup> adj	0.113 / 0.106			0.308 / 0.296			0.201 / 0.192			0.383 / 0.372		
AIC	-486.455			-566.538			-520.519			-606.372		

**Remote and unsupervised short assessments show moderate-to-good retest reliability**

Finally, we were interested in retest reliability of remote and unsupervised assessments given that unsupervised and remote memory assessments are well suited to assess longitudinal memory trajectories. To that end we limited the dataset to individuals that had at least four sessions completed and calculated two-way random intraclass correlation coefficients (ICC) for two scenarios. Regarding a scenario where single test sessions would be used, we calculated the ICC between test session 1 and 2 and regarding a scenario where the average between two sessions would be used, we calculated the ICC between the average of test session 1 and 2 and the average of test session 3 and 4. A single session of MDT-OS (ICC of 0.64, 95% CI [0.48, 0.76]; MDT-O: 0.57 [CI 0.39, 0.71]; MDT-S: 0.49 [CI 0.29, 0.65]; n=71) and a single session of ORR (ICC of 0.64 [CI 0.4, 0.8]; n=38) both showed moderate retest reliability. In the second scenario, the average of two test sessions of MDT-OS (ICC of 0.79 [CI 0.68, 0.87]; MDT-O: 0.64 [CI 0.48, 0.76]; MDT-S: 0.62 [CI 0.41, 0.76]), as well as the average of two test sessions of ORR (ICC of 0.81 [CI 0.67, 0.9]) showed good retest reliability.

## **Discussion**

We found construct validity of remote and unsupervised digital memory assessments to be good when comparing them with in-clinic traditional neuropsychological assessments, and that baseline performance in the remote memory assessments was associated with future decline in the mPACC score. The model that best predicted future cognitive decline in the mPACC included a combination of plasma p-tau217 and MDT-S. In a restricted sample of individuals where both digital markers were completed, the model that best predicted future cognitive decline in the mPACC also included the ORR-DR. Retest reliability was moderate-to-good when repeating the tests utilizing parallel test versions. In addition, digital memory assessments were significantly associated with PET and CSF measures of tau pathology as well as downstream MRI measures of MTL atrophy. Finally, the onboarding to the study as well as the unsupervised completion of memory tasks itself was rated positively by participants via questionnaires.

### **Older adults and patients were able to complete unsupervised digital assessments**

There exist stereotypes regarding older adults' and patients' familiarity with smartphones and tablet computers and their willingness to participate in remote and unsupervised studies using mobile devices. However, recent work has shown that, while older adults were indeed less familiar with technology, the majority of older participants decided to participate in remote studies and showed exceptional adherence when studies were planned thoughtful and included user-centered design (Nicosia et al., 2022a). In line with this, we found that a large majority of elderly participants owned a smartphone or tablet and downloaded apps by themselves or with help. The participants rated both of the current remote memory assessments as challenging but not too difficult. However, while it was intended that participants perform the ORR with a delay period of 60 minutes, there was high variability of delay times due to participants postponing the recall tasks for too long. Given that the delay period is a critical factor for task difficulty, we thus excluded all sessions with delay periods longer than 4 hours which resulted in a mean delay period of 96 minutes. Future studies should incorporate appropriate counter measures to minimize this delay period.

The main reason to not participate in the smartphone study was the high frequency of test sessions. However, the sub-sample of participants who finished the study after one year found that the number of tests was just right. A large majority found the instructions clear and the application easy to use. Three-quarters even indicated they would prefer remote mobile cognitive tests over traditional in person paper-pencil assessments. Overall, the app was rated seven out of ten by more than 80% of participants. However, participants that dropped out of the study rated the app lower and found that the number of tests were too high and durations were too long. Overall this indicates that

the app was usable and easy to work with for our study participants, but that the intense continuous study schedule might be optimized.

### **Remote and unsupervised assessments reflect traditional neuropsychological assessments**

When introducing a novel cognitive measure, its construct validity needs to be assessed by comparing it against established neuropsychological measures of constructs it is intended to measure. This is particularly true for unsupervised and remote digital assessments that are completed in an unstandardized environment of the participant's choice. Recent work using smartphone-based cognitive assessments in samples of older participants could successfully demonstrate high construct validity across various cognitive domains (Nicosia et al., 2022b; Öhman et al., 2022; Papp et al., 2021; Singh et al., 2023; Vyshedskiy et al., 2022). Both the ORR and the MDT-OS are intended to measure memory. While the ORR specifically aims to measure delayed memory, the MDT-OS aims to measure precision memory in an n-back task design known to also rely on attentional and executive processes (Gajewski et al., 2018). Accordingly, we found that while the ORR-DR was most strongly associated with ADAS delayed recall and less with SDMT, verbal fluency and the MMSE, the MDT-OS was associated with SDMT, ADAS delayed recall and verbal fluency and less so with the MMSE. The strongest relationship for both measures, however, was found with the mPACC. These findings nicely support earlier work using the neotiv memory tasks in samples with cognitively unimpaired individuals and MCI patients (Berron et al., 2023; Öhman et al., 2022). Furthermore, baseline performance in the MDT-S and the ORR-DR was associated with future decline in the mPACC in this study, demonstrating first evidence of its prognostic validity regarding future cognitive decline.

While high construct validity indicates that the ORR and the MDT-OS were successfully designed to measure memory function and in part executive and attentional processes, it also indicates that study participants were seemingly successful in completing the cognitive tests in an appropriate environment. While we do not know details about the individual test environment, our findings indicate that it was generally an undisturbed environment as 88% of all test sessions were reported to be without any distractions. Interestingly, our findings show that cognitively unimpaired individuals reported more distractions compared to MCI patients, in line with earlier research (Madero et al., 2021).

### **Remote and unsupervised assessments are sensitive to measures of AD pathology**

Both the MDT-OS and the ORR are considered MTL-dependent tasks (Berron et al., 2019, 2018; Grande et al., 2022, 2021; Maass et al., 2019) and earlier work has already linked performance in the MDT-OS with fluid and imaging measures of AD pathology (Berron et al., 2019; Maass et al., 2019). In line with these earlier findings, we found that MDT-O and ORR-DR were both significantly associated

with tau-PET measures from the hippocampus and area 35, two brain regions that are affected by tau pathology very early on in the course of AD (Berron et al., 2021). Also, in line with earlier work, we did not observe such a relationship for the MDT-S (Berron et al., 2019). Finally, we found that remote digital assessments were associated with MTL atrophy, which is considered a downstream effect of AD pathology. Area 35 and the hippocampus are among the earliest sites of atrophy in AD (Berron et al., 2021; Xie et al., 2020, 2018). While the MDT-OS was associated with atrophy of the hippocampus and area 35, ORR-DR was only associated with posterior hippocampal atrophy. Taken together, this shows that both tasks depend on the functional integrity of the MTL and task performance is affected by early accumulation of AD pathology.

### **Potential of remote assessments for case finding, monitoring and prognosis of cognitive impairment in Alzheimer's disease**

Remote and unsupervised digital cognitive assessments hold promise for case finding in health care and clinical trials, but also for longitudinal cognitive monitoring and even prognosis of cognitive decline and disease progression. Recent work showed that outcomes from remote assessments can help to identify MCI patients (Berron et al., 2023) and potentially even A $\beta$ -positive but cognitively unimpaired participants (Jutten et al., 2022; Samaroo et al., 2020). Regarding prognosis, recent work has shown that plasma biomarkers for tau pathology are associated with future cognitive decline (Mattsson-Carlgren et al., 2023), and in combination with brief in-clinic cognitive tests, can identify MCI patients who are likely to progress towards dementia (Palmqvist et al., 2021). High construct validity and first evidence of prognostic validity regarding cognitive decline in the mPACC suggest that remote memory assessments may also have potential utility in prognosis. Indeed, our analysis comparing models with plasma p-tau217 against a model combining plasma p-tau217 and remote digital memory assessments suggests that incorporating remote digital memory assessments can significantly contribute to the prediction of future decline. Moreover, the moderate-to-good retest reliability of both remote memory assessments, coupled with the minimal practice effects due to the utilization of parallel test sets (Berron et al., 2022) hints that these tests might prove beneficial in monitoring cognitive change over time. While retest-reliability was moderate when using one single assessment, it was good when averaging across only two sessions. Future studies should investigate whether longitudinal trajectories derived from remote and unsupervised cognitive assessments can capture subtle cognitive decline.

We need to carefully consider some limitations to this study. Firstly, the sample size is relatively modest, particularly for participants who completed both digital remote and unsupervised memory paradigms. Secondly, our implementation of the ORR did not strictly enforce adherence to the planned

retrieval delay intervals, which led some individuals to perform recall assessments after prolonged delays. Given the impact of delay length on task performance, we excluded sessions with significantly extended delay periods (more than 240 minutes) resulting in substantial reduction of test sessions in this study. Thus, future implementations of this task, and remote and unsupervised assessments of long-term memory in general, should make it easier for participants and patients to integrate remote and repeated tests into their everyday life, while still enforcing minimized delay periods.

## **Conclusion**

Our results demonstrate that unsupervised and remote digital memory assessments could effectively become a valuable tool in the diagnosis and prognosis of AD, conceivably in combination with plasma biomarkers.

## References

- Baker, S.L., Maass, A., Jagust, W.J., 2017. Considerations and code for partial volume correcting [18F]-AV-1451 tau PET data. *Data Brief* 15, 648–657. <https://doi.org/10.1016/j.dib.2017.10.024>
- Bakker, A., Kirwan, B.C., Miller, M., Stark, C.E., 2008. Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus. *Science* 319, 1640–1642. <https://doi.org/10.1126/science.1152882>
- Berron, D., Cardenas-Blanco, A., Bittner, D., Metzger, C.D., Spottke, A., Heneka, M.T., Fließbach, K., Schneider, A., Teipel, S.J., Wagner, M., Speck, O., Jessen, F., Düzel, E., 2019. Higher CSF tau levels are related to hippocampal hyperactivity and object mnemonic discrimination in older adults. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. <https://doi.org/10.1523/jneurosci.1279-19.2019>
- Berron, D., Glanz, W., Clark, L., Basche, K., Grande, X., Güsten, J., Billette, O.V., Hempen, I., Naveed, M.H., Diersch, N., Butryn, M., Spottke, A., Buerger, K., Perneczky, R., Schneider, A., Teipel, S., Wiltfang, J., Johnson, S., Wagner, M., Jessen, F., Düzel, E., Consortium, the D., 2023. A Remote Digital Memory Composite to Detect Cognitive Impairment in Memory Clinic Samples in Unsupervised Settings using Mobile Devices. *medRxiv* 2021.11.12.21266226. <https://doi.org/10.1101/2021.11.12.21266226>
- Berron, D., Neumann, K., Maass, A., Schütze, H., Fließbach, K., Kiven, V., Jessen, F., Sauvage, M., Kumaran, D., Düzel, E., 2018. Age-related functional changes in domain-specific medial temporal lobe pathways. *Neurobiology of aging* 65, 86–97. <https://doi.org/10.1016/j.neurobiolaging.2017.12.030>
- Berron, D., Schütze, H., Maass, A., Cardenas-Blanco, A., Kuijf, H.J., Kumaran, D., Düzel, E., 2016. Strong Evidence for Pattern Separation in Human Dentate Gyrus. *The Journal of Neuroscience* 36, 7569–7579. <https://doi.org/10.1523/jneurosci.0518-16.2016>
- Berron, D., Vogel, J.W., Insel, P.S., Pereira, J.B., Xie, L., Wisse, L.E.M., Yushkevich, P.A., Palmqvist, S., Mattsson-Carlsson, N., Stomrud, E., Smith, R., Strandberg, O., Hansson, O., 2021. Early stages of tau pathology and its associations with functional connectivity, atrophy and memory. *Brain* 144, awab114-. <https://doi.org/10.1093/brain/awab114>
- Berron, D., Ziegler, G., Vieweg, P., Billette, O., Güsten, J., Grande, X., Heneka, M.T., Schneider, A., Teipel, S., Jessen, F., Wagner, M., Düzel, E., 2022. Feasibility of Digital Memory Assessments in an Unsupervised and Remote Study Setting. *Frontiers Digital Heal* 4, 892997. <https://doi.org/10.3389/fdgth.2022.892997>
- Burnham, K.P., Anderson, D.R., 2004. Multimodel Inference. *Sociol. Methods Res.* 33, 261–304. <https://doi.org/10.1177/0049124104268644>
- Cho, S.H., Choe, Y.S., Park, S., Kim, Y.J., Kim, H.J., Jang, H., Kim, S.J., Kim, J.P., Jung, Y.H., Kim, B.C., Na, D.L., Moon, S.H., Seo, S.W., 2020. Appropriate reference region selection of 18F-florbetaben and 18F-flutemetamol beta-amyloid PET expressed in Centiloid. *Sci Rep-uk* 10, 14950. <https://doi.org/10.1038/s41598-020-70978-z>
- Flores, R. de, Das, S.R., Xie, L., Wisse, L.E.M., Lyu, X., Shah, P., Yushkevich, P.A., Wolk, D.A., 2022. Medial Temporal Lobe Networks in Alzheimer’s Disease: Structural and Molecular Vulnerabilities. *J Neurosci* 42, 2131–2141. <https://doi.org/10.1523/jneurosci.0949-21.2021>
- Gajewski, P.D., Hanisch, E., Falkenstein, M., Thönes, S., Wascher, E., 2018. What Does the n-Back Task Measure as We Get Older? Relations Between Working-Memory Measures and Other Cognitive Functions Across the Lifespan. *Front Psychol* 9, 2208. <https://doi.org/10.3389/fpsyg.2018.02208>
- Grande, X., Berron, D., Horner, A.J., Bisby, J.A., Düzel, E., Burgess, N., 2019. Holistic recollection via pattern completion involves hippocampal subfield CA3. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. <https://doi.org/10.1523/jneurosci.0722-19.2019>
- Grande, X., Berron, D., Maass, A., Bainbridge, W., Düzel, E., 2021. Content-specific vulnerability of recent episodic memories in Alzheimer’s disease. *Neuropsychologia* 107976. <https://doi.org/10.1016/j.neuropsychologia.2021.107976>
- Grande, X., Sauvage, M.M., Becke, A., Düzel, E., Berron, D., 2022. Transversal functional connectivity and scene-specific processing in the human entorhinal-hippocampal circuitry. *Elife* 11. <https://doi.org/10.7554/elife.76479>
- Grothe, M.J., Barthel, H., Sepulcre, J., Dyrba, M., Sabri, O., Teipel, S.J., Initiative, A., 2017. In vivo staging of regional amyloid deposition. *Neurology* 89, 2031–2038. <https://doi.org/10.1212/wnl.0000000000004643>
- Hansson, O., 2021. Biomarkers for neurodegenerative diseases. *Nat Med* 1–10. <https://doi.org/10.1038/s41591-021-01382-x>
- Hansson, O., Blennow, K., Zetterberg, H., Dage, J., 2023. Blood biomarkers for Alzheimer’s disease in clinical practice and trials. *Nat. Aging* 3, 506–519. <https://doi.org/10.1038/s43587-023-00403-3>
- Jutten, R.J., Rentz, D.M., Fu, J.F., Mayblyum, D.V., Amariglio, R.E., Buckley, R.F., Properzi, M.J., Maruff, P., Stark, C.E., Yassa, M.A., Johnson, K.A., Sperling, R.A., Papp, K.V., 2022. Monthly At-Home Computerized Cognitive Testing to Detect Diminished Practice Effects in Preclinical Alzheimer’s Disease. *Front Aging Neurosci* 13, 800126. <https://doi.org/10.3389/fnagi.2021.800126>

- Koo, T.K., Li, M.Y., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* 15, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Maass, A., Berron, D., Harrison, T.M., Adams, J.N., Joie, R.L., Baker, S., Mellinger, T., Bell, R.K., Swinnerton, K., Inglis, B., Rabinovici, G.D., Düzel, E., Jagust, W.J., 2019. Alzheimer's pathology targets distinct memory networks in the ageing brain. *Brain : a journal of neurology* 142, 2492–2509. <https://doi.org/10.1093/brain/awz154>
- Maass, A., Berron, D., Libby, L.A., Ranganath, C., Düzel, E., 2015. Functional subregions of the human entorhinal cortex. *Elife* 4, e06426. <https://doi.org/10.7554/elife.06426>
- Madero, E.N., Anderson, J., Bott, N.T., Hall, A., Newton, D., Fuseya, N., Harrison, J.E., Myers, J.R., Glenn, J.M., 2021. Environmental Distractions during Unsupervised Remote Digital Cognitive Assessment. *J Prev Alzheimer's Dis* 8, 263–266. <https://doi.org/10.14283/jpad.2021.9>
- Mattsson-Carlgrén, N., Salvadó, G., Ashton, N.J., Tideman, P., Stomrud, E., Zetterberg, H., Ossenkoppele, R., Betthausen, T.J., Cody, K.A., Jonaitis, E.M., Langhough, R., Palmqvist, S., Blennow, K., Janelidze, S., Johnson, S.C., Hansson, O., 2023. Prediction of Longitudinal Cognitive Decline in Preclinical Alzheimer Disease Using Plasma Biomarkers. *Jama Neurol* 80. <https://doi.org/10.1001/jamaneurol.2022.5272>
- Nicosia, J., Aschenbrenner, A.J., Adams, S.L., Tahan, M., Stout, S.H., Wilks, H., Balls-Berry, J.E., Morris, J.C., Hassenstab, J., 2022a. Bridging the Technological Divide: Stigmas and Challenges With Technology in Digital Brain Health Studies of Older Adults. *Frontiers Digital Heal* 4, 880055. <https://doi.org/10.3389/fdgth.2022.880055>
- Nicosia, J., Aschenbrenner, A.J., Balota, D., Sliwinski, M., Tahan, M., Adams, S., Stout, S.H., Wilks, H., Gordon, B., Benzinger, T., Fagan, A., Xiong, C., Bateman, R., Morris, J.C., Hassenstab, J., n.d. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer disease. <https://doi.org/10.31234/osf.io/wtsyn>
- Nicosia, J., Aschenbrenner, A.J., Balota, D.A., Sliwinski, M.J., Tahan, M., Adams, S., Stout, S.S., Wilks, H., Gordon, B.A., Benzinger, T.L.S., Fagan, A.M., Xiong, C., Bateman, R.J., Morris, J.C., Hassenstab, J., 2022b. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. *J Int Neuropsych Soc* 29, 459–471. <https://doi.org/10.1017/s135561772200042x>
- Öhman, F., Berron, D., Papp, K.V., Kern, S., Skoog, J., Bodin, T.H., Zettergren, A., Skoog, I., Schöll, M., 2022. Unsupervised mobile app-based cognitive testing in a population-based study of older adults born 1944. *Frontiers Digital Heal* 4, 933265. <https://doi.org/10.3389/fdgth.2022.933265>
- Öhman, F., Hassenstab, J., Berron, D., Schöll, M., Papp, K.V., 2021. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimer's Dementia Diagnosis Assess Dis Monit* 13, e12217. <https://doi.org/10.1002/dad2.12217>
- Palmqvist, S., Janelidze, S., Quiroz, Y.T., Zetterberg, H., Lopera, F., Stomrud, E., Su, Y., Chen, Y., Serrano, G.E., Leuzy, A., Mattsson-Carlgrén, N., Strandberg, O., Smith, R., Villegas, A., Sepulveda-Falla, D., Chai, X., Proctor, N.K., Beach, T.G., Blennow, K., Dage, J.L., Reiman, E.M., Hansson, O., 2020. Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. *Jama* 324, 772–781. <https://doi.org/10.1001/jama.2020.12134>
- Palmqvist, S., Schöll, M., Strandberg, O., Mattsson, N., Stomrud, E., Zetterberg, H., Blennow, K., Landau, S., Jagust, W., Hansson, O., 2017. Earliest accumulation of  $\beta$ -amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nature Communications* 8, 1214. <https://doi.org/10.1038/s41467-017-01150-x>
- Palmqvist, S., Tideman, P., Cullen, N., Zetterberg, H., Blennow, K., Dage, J.L., Stomrud, E., Janelidze, S., Mattsson-Carlgrén, N., Hansson, O., 2021. Prediction of future Alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. *Nat Med* 1–9. <https://doi.org/10.1038/s41591-021-01348-z>
- Papp, K.V., Samaroo, A., Chou, H., Buckley, R., Schneider, O.R., Hsieh, S., Soberanes, D., Quiroz, Y., Properzi, M., Schultz, A., García-Magariño, I., Marshall, G.A., Burke, J.G., Kumar, R., Snyder, N., Johnson, K., Rentz, D.M., Sperling, R.A., Amariglio, R.E., 2021. Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimer's Dementia Diagnosis Assess Dis Monit* 13. <https://doi.org/10.1002/dad2.12243>
- Quadalti, C., Palmqvist, S., Hall, S., Rossi, M., Mammana, A., Janelidze, S., Dellavalle, S., Mattsson-Carlgrén, N., Baiardi, S., Stomrud, E., Hansson, O., Parchi, P., 2023. Clinical effects of Lewy body pathology in cognitively impaired individuals. *Nat. Med.* 1–7. <https://doi.org/10.1038/s41591-023-02449-7>
- Ranganath, C., Ritchey, M., 2012. Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci* 13, 713–26. <https://doi.org/10.1038/nrn3338>
- Ritchey, M., Cooper, R.A., 2020. Deconstructing the Posterior Medial Episodic Network. *Trends Cogn Sci.* <https://doi.org/10.1016/j.tics.2020.03.006>
- Rousset, O.G., Ma, Y., Evans, A.C., 1998. Correction for partial volume effects in PET: principle and validation. *J Nucl Medicine Official Publ Soc Nucl Medicine* 39, 904–11.



- Samaroo, A., Amariglio, R.E., Burnham, S., Sparks, P., Properzi, M., Schultz, A.P., Buckley, R., Johnson, K.A., Sperling, R.A., Rentz, D.M., Papp, K.V., 2020. Diminished Learning Over Repeated Exposures (LORE) in preclinical Alzheimer's disease. *Alzheimer's Dementia Diagnosis Assess Dis Monit* 12, e12132. <https://doi.org/10.1002/dad2.12132>
- Sanchez, J.S., Becker, J.A., Jacobs, H.I.L., Hanseeuw, B.J., Jiang, S., Schultz, A.P., Properzi, M.J., Katz, S.R., Beiser, A., Satizabal, C.L., O'Donnell, A., DeCarli, C., Killiany, R., Fakhri, G.E., Normandin, M.D., Gómez-Isla, T., Quiroz, Y.T., Rentz, D.M., Sperling, R.A., Seshadri, S., Augustinack, J., Price, J.C., Johnson, K.A., 2021. The cortical origin and initial spread of medial temporal tauopathy in Alzheimer's disease assessed with positron emission tomography. *Sci Transl Med* 13. <https://doi.org/10.1126/scitranslmed.abc0655>
- Schröder, T., Haak, K.V., Jimenez, N.I., Beckmann, C.F., Doeller, C.F., 2015. Functional topography of the human entorhinal cortex. *eLife* 4, e06738. <https://doi.org/10.7554/elife.06738>
- Singh, S., Strong, R., Xu, I., Fonseca, L.M., Hawks, Z., Grinspoon, E., Jung, L., Li, F., Weinstock, R.S., Sliwinski, M.J., Chaytor, N.S., Germine, L.T., 2023. Ecological Momentary Assessment of Cognition in Clinical and Community Samples: Reliability and Validity Study. *J Med Internet Res* 25, e45028. <https://doi.org/10.2196/45028>
- Thomas, B.A., Erlandsson, K., Modat, M., Thurfjell, L., Vandenberghe, R., Ourselin, S., Hutton, B.F., 2011. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur J Nucl Med Mol I* 38, 1104–1119. <https://doi.org/10.1007/s00259-011-1745-9>
- Thurfjell, L., Lilja, J., Lundqvist, R., Buckley, C., Smith, A., Vandenberghe, R., Sherwin, P., 2014. Automated Quantification of 18F-Flutemetamol PET Activity for Categorizing Scans as Negative or Positive for Brain Amyloid: Concordance with Visual Image Reads. *J Nucl Med* 55, 1623–1628. <https://doi.org/10.2967/jnumed.114.142109>
- Vyshedskiy, A., Netson, R., Fridberg, E., Jagadeesan, P., Arnold, M., Barnett, S., Gondalia, A., Maslova, V., Torres, L. de, Ostrovsky, S., Durakovic, D., Savchenko, A., McNett, S., Kogan, M., Piryatinsky, I., Gold, D., 2022. Boston cognitive assessment (BOCA) — a comprehensive self-administered smartphone- and computer-based at-home test for longitudinal tracking of cognitive performance. *Bmc Neurol* 22, 92. <https://doi.org/10.1186/s12883-022-02620-6>
- Xie, L., Das, S.R., Wisse, L., Ittyerah, R., Yushkevich, P.A., Wolk, D.A., Initiative, A., 2018. Early Tau Burden Correlates with Higher Rate of Atrophy in Transentorhinal Cortex. *Journal of Alzheimer's disease : JAD* 62, 85–92. <https://doi.org/10.3233/jad-170945>
- Xie, L., Pluta, J.B., Das, S.R., Wisse, L.E., Wang, H., Mancuso, L., Kliot, D., Avants, B.B., Ding, S.-L., Manjón, J.V., 2017. Multi-template analysis of human perirhinal cortex in brain MRI: explicitly accounting for anatomical variability. *NeuroImage* 144, 183–202.
- Xie, L., Wisse, L.E.M., Das, S.R., Vergnet, N., Dong, M., Ittyerah, R., Flores, R., Yushkevich, P.A., Wolk, D.A., Initiative, for the A.D.N., 2020. Longitudinal atrophy in early Braak regions in preclinical Alzheimer's disease. *Hum Brain Mapp*. <https://doi.org/10.1002/hbm.25151>
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

## **Acknowledgements**

We want to thank all participants of the Swedish BioFINDER-2 study and their families for their participation in the study. We also want to thank Oisín Clancy and Hannah Baumeister for their help with the manual quality assessment of the medial temporal subregional masks.

## **Funding**

Work at the authors' research center was supported by the Alzheimer's Association (SG-23-1061717), Swedish Research Council (2022-00775, 2018-02052), ERA PerMed (ERAPERMED2021-184), the Knut and Alice Wallenberg foundation (2017-0383), the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation (AF-980907, AF-981132), the Swedish Brain Foundation (FO2021-0293, FO2022-0204), The Parkinson foundation of Sweden (1412/22), the Cure Alzheimer's fund, the Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, the Skåne University Hospital Foundation (2020-0000028), Regionalt Forskningsstöd (2022-1259) and the Swedish federal government under the ALF agreement (2022-Projekt0080). DB was supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 843074 and the donors of Alzheimer's Disease Research, a program of the BrightFocus Foundation. The precursor of <sup>18</sup>F-flutemetamol was sponsored by GE Healthcare. The precursor of <sup>18</sup>F-RO948 was provided by Roche. The funding sources had no role in the design and conduct of the study; in the collection, analysis, interpretation of the data; or in the preparation, review, or approval of the manuscript.

## **Competing interests**

OH has acquired research support (for the institution) from ADx, AVID Radiopharmaceuticals, Biogen, Eli Lilly, Eisai, Fujirebio, GE Healthcare, Pfizer, and Roche. In the past 2 years, he has received consultancy/speaker fees from AC Immune, Amylyx, Alzpath, BioArctic, Biogen, Cerveau, Eisai, Eli Lilly, Fujirebio, Merck, Novartis, Novo Nordisk, Roche, Sanofi and Siemens. SP has acquired research support (for the institution) from ki elements/ADDF and Avid. In the past 2 years, he has received consultancy/speaker fees from Bioarctic, Biogen, Eisai, Lilly, and Roche. ED reports personal fees from Biogen, Roche, Lilly, Eisai and UCL Consultancy as well as non-financial support from Rox Health. DB and ED are scientific co-founders of neotiv GmbH and own company shares. The other authors report no competing interests.