Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

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

David Berron[1,2] | Emil Olsson[1] | Felix Andersson[3] | Shorena Janelidze[1] |
Pontus Tideman[1,3] | Emrah Düzel[2,4,5] | Sebastian Palmqvist[1,3] | Erik Stomrud[1,3] |
Oskar Hansson[1,3]

[1]Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Lund, Sweden

[2]German Center for Neurodegenerative Diseases, Magdeburg, Germany

[3]Memory Clinic, Skåne University Hospital, Malmö, Sweden

[4]Institute for Cognitive Neurology and Dementia Research, Otto-von-Guericke University, Magdeburg, Germany

[5]Institute of Cognitive Neuroscience, University College London, London, UK

Correspondence
David Berron and Oskar Hansson, Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Sölvegatan 19, 223 62 Lund, Sweden.
Email: david.berron@dzne.de, oskar.hansson@med.lu.se

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-positron emission tomography and downstream magnetic resonance imaging 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, ambulatory assessments, blood-based biomarkers, digital cognitive markers, ecological momentary assessments, memory, mHealth, plasma marker, smartphone-based unsupervised assessments

Highlights

- Remote and unsupervised digital memory assessments are feasible in older adults and individuals in early stages of Alzheimer's disease.
- Digital memory assessments are associated with neuropsychological in-clinic assessments, tau-positron emission tomography and magnetic resonance imaging measures.
- Combination of digital memory assessments with plasma p-tau217 holds promise for prognosis of future cognitive decline.
- Future validation in further independent, larger, and more diverse cohorts is needed to inform clinical implementation.

1 | BACKGROUND

While there has been significant progress in fluid and neuroimaging biomarkers to detect pathological changes in Alzheimer's disease (AD), most cognitive measures that are used in health care and clinical trials were initially designed to detect overt cognitive impairment and novel developments still lag behind.[1,2] 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.[3,4] Recent work on the functional neuroanatomy of episodic memory has 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.[4–7] Episodic memory requires pattern separation and completion processes that are primarily mediated by MTL subregions. While the dentate gyrus is involved in pattern separation[8,9] and reduces memory interference between similar events, the hippocampal Cornu Ammonis 3 mediates pattern completion processes in interplay with neocortical regions.[10] 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.[3,11–15] 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.[16]

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 are gradually implementing decentralized clinical trial structures for case-finding and monitoring,[17] for example, in 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.[18] Approaches from several groups, including our own, have recently demonstrated that remote and unsupervised assessments are feasible in healthy older adult populations and those at risk of AD.[19–24] Furthermore, this work showed that remote and unsupervised assessments can support the identification of mild cognitive impairment (MCI) patients in a memory clinic setting[25] and potentially even β-amyloid (Aβ) positive but cognitively unimpaired (CU) participants.[19,26,27] 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, and their relationship with fluid and imaging biomarkers of AD. To that end, we implemented two non-verbal visual memory tasks based on recent insights into the functional anatomy of episodic memory, available on the neotiv digital platform (https://www.neotiv.com/en),[21,24] in a subset of the Swedish BioFINDER-2 study.

2 | METHODS

2.1 | Recruitment into smartphone-based add-on study

A total of 187 individuals gave written informed consent to participate in the 7 Tesla magnetic resonance imaging (MRI) study arm of the Swedish BioFINDER2 study and were offered participation in biweekly smartphone-based remote cognitive assessments across a 12-month period. Of those offered participation, 160 individuals agreed to complete the smartphone-based assessments. 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 on their own mobile device 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 who 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 but could also be postponed. Each of the assessments consisted of a two-phase session. The two phases were either two halves of a mnemonic discrimination task (MDT-OS) or encoding and retrieval phases of an object-in-room recall (ORR) task (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. To minimize potential practice effects from repeated testing, 12 independent difficulty-matched parallel test versions for each memory task were used.[24] 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). In the current cross-sectional analyses, the first valid session of each participant was used.

Push notifications were used to notify individuals about available tasks and to remind them daily for five consecutive days if a task 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
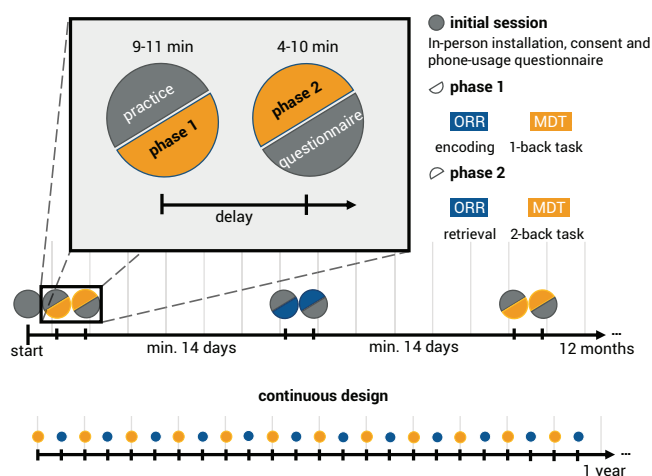
FIGURE 1 Timeline of the study protocol. Participants enlisted for a 12-month study of biweekly remote and unsupervised memory assessments of the MDT-OS and ORR. In the initial in-clinic session, they gave consent, got a brief introduction, answered a short questionnaire on their phone usage, and completed the first task. Every 2nd 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. MDT-OS; Mnemonic Discrimination Task for Objects and Scenes; ORR, objects-in-room-recall task.

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, 38 participants completed telephone interviews to provide feedback on usability and user experience throughout the study. The full study timeline can be seen in Figure 1.

2.2 | Smartphone-based memory assessments

2.2.1 | Mnemonic discrimination of objects and scenes

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 Figure 2A). 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-h 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 Mnemonic Discrimination of Objects and Scenes (MDT-OS) has been designed to tax hippocampal pattern separation; a memory

(A)

Mnemonic Discrimination for Objects and Scenes (MDT-OS)
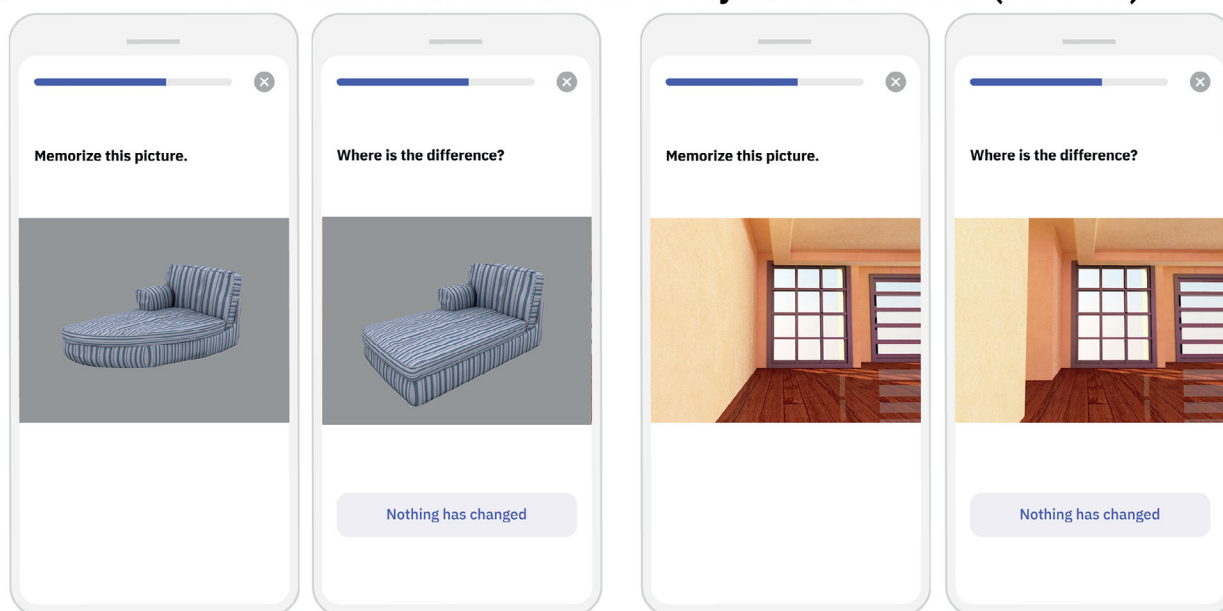
(B)

Object-in-Room Recall (ORR)
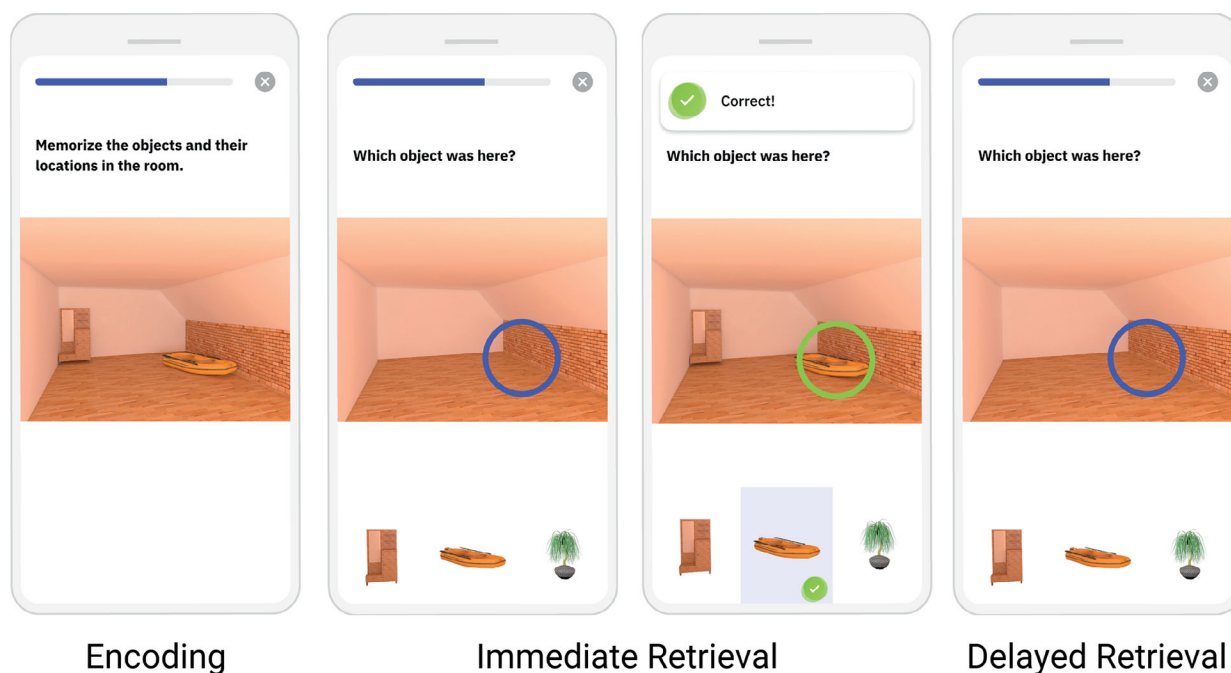
Encoding   Immediate Retrieval   Delayed Retrieval

FIGURE 2 Smartphone-based memory tasks. (A) MDT-OS and (B) ORR test. MDT-OS; Mnemonic Discrimination Test for Objects and Scenes; ORR, Objects-In-Room Recall.

mechanism needed to discriminate between similar memories. Earlier studies using functional MRI have shown that especially subregions in the human MTL are involved in this task.[3,11,13] The test provides a hit rate, a false alarm rate and a corrected hit rate (hit rate minus false alarm rate) for both the object and scene condition. The averaged corrected hit rate across the scene and the object condition was used as the main outcome measure in the following analyses and object, and scene-specific corrected hit rates were additionally investigated.

2.2.2 | Objects-In-Room Recall

In this task, participants had to memorize a spatial arrangement of 2 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 3 possible objects, 1 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.[10,28] 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 (ORR-DR [delayed recall]; 0–25) score. Here, we use the ORR-DR.

2.2.3 | Quality control procedures

In the current analyses, we analyzed data collected up until the data release in July 2023. Recruitment of participants from the Swedish BioFINDER-2 study was done between February 2019 and February 2022. Only participants who had completed at least 1 full session, that is, both phases of a task, were included. Regarding the 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, 81% of test sessions we report here were from the first 2 MDT-OS and ORR-DR sessions respectively while 19% were from subsequent test sessions (mean test session = 2.1, range = 1–11).

2.3 | Cerebrospinal fluid and plasma analysis

Cerebrospinal fluid (CSF) Aβ42 and Aβ40 concentrations were measured using the Elecsys assays (Roche Diagnostics). Participants were stratified for Aβ status using the CSF Aβ42/40 ratio with a predefined cut-off of 0.08.[29] Plasma levels of p-tau217 were quantified with the MesoScale Discovery-based immunoassay developed by Lilly Research Laboratories as previously described.[30]

2.4 | Imaging acquisition

2.4.1 | MRI

T1-weighted images were acquired on a 3T Siemens Prisma scanner (Siemens Medical Solutions, Erlangen, Germany) with a 64-channel head coil using an MPRAGE sequence (in-plane resolution = 1 × 1 mm^2, slice thickness = 1 mm, TR = 1900 ms, echo time = 2.54 ms, flip-angle = 9°). The 7T data acquired in the project were not analyzed for this manuscript.

2.4.2 | Tau and Aβ-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 [18F]RO948, and LIST mode emission data was acquired for each scan 70–90 minutes ([18F]RO948) post injection. Aβ-PET imaging was performed on the same platform 90–110 minutes after the injection of ∼185 MBq [18F]flutemetamol.[31]

2.5 | Imaging analysis

2.5.1 | 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 (1 × 1 × 1 mm^3 resolution) using Automatic Segmentation of Hippocampal Subfields-T1 (Xie et al., 2019) and a multi-template thickness analysis pipeline.[32] All subregional masks were visually assessed.

2.5.2 | Standardized uptake value ratio measures

[18F]RO948: Standardized uptake value ratio (SUVr) images were calculated for area 35 using an inferior cerebellar reference region.[33] Partial volume correction was performed using the geometric transfer matrix method[34] extended to voxel-level using region-based voxel-wise correction.[35] To reduce the influence of off-target binding, choroid plexus tau-PET signal was regressed from hippocampal measures. [18F]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.[36]

2.5.3 | 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 (Rosen et al., 1984). The learning trial of the 10 words was repeated 3 times. After a distraction task (Boston Naming—15 items short version[37]), 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

TABLE 1 Participant demographics of the entire sample.

Parameter	CU Aβ− (N = 49)	CU Aβ+ (N = 28)	MCI Aβ− (N = 9)	MCI Aβ+ (N = 14)	Total (N = 100)
Age (years)	60.6 (10.9)	69.8 (8.4)	64 (11.9)	67.8 (4.3)	64.5 (10.4)
Education (years)	13.5 (3.3)	12.9 (3.3)	12.7 (2.7)	12.5 (3.8)	13.1 (3.3)
Sex (% female)	53.1%	57.1%	55.6%	35.7%	52%
MMSE	29 (1.2)	28.6 (1.6)	27 (2.1)	27.7 (1.5)	28.5 (1.6)

Note: Displayed are mean values (standard deviations) unless otherwise stated.
Abbreviations: CU, cognitively unimpaired; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; *N*, number of participants.

cognition, we used the Mini-Mental State Examination (MMSE[38]; scale from 0–30). For attentional and executive function, we used the Symbol Digit Modalities Test (SDMT,[39] 1 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 during a functional MRI scan. In the end, we calculated a composite score similar to the Preclinical Alzheimer Cognitive Composite 5 (PACC5[40]), the modified PACC (mPACC), using the average of the z-standardized scores of Alzheimer's Disease Assessment Scale (ADAS) delayed recall (counted twice), SDMT, MMSE, and verbal fluency. The average time between the mPACC and the remote digital assessments was 182 days.

2.5.4 | Statistical analysis

Multiple regression analyses were carried out between PET measures, MTL subregional atrophy and unsupervised digital remote memory tests in R (version 4.1.2; 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 (on average 3.4 mPACC timepoints per participant across up to 5 years). 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 the Akaike information criterion (AIC), where a difference of 2 or greater is considered significant.[41] Retest reliability was assessed for 2 individual sessions (first and second session of the respective test; on average after 7 weeks) and for 2 averaged sessions (mean of first and second session of the respective test vs. mean of third and fourth session of the respective test; on average after 13 weeks). Intraclass correlation coefficients (ICCs) and their 95%

confident intervals were calculated based on single rater, absolute-agreement, 2-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.[42]
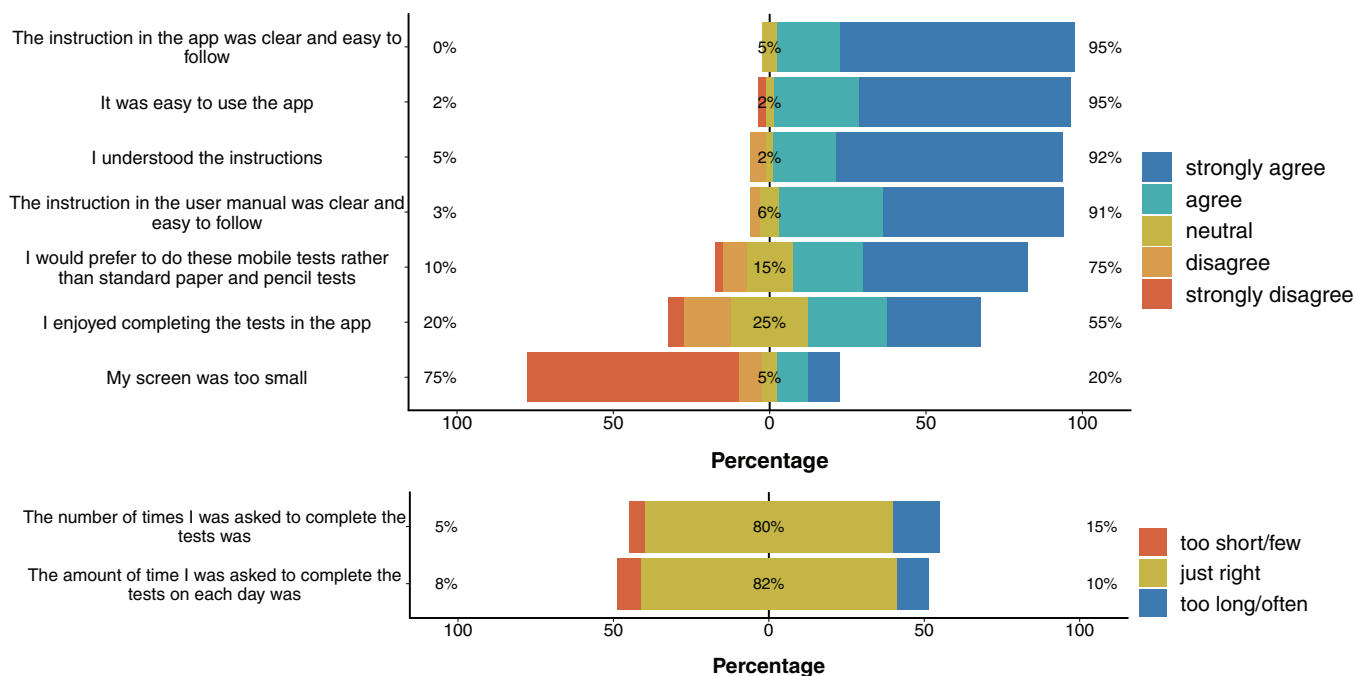
3 | RESULTS

3.1 | Participant sample

Here, we considered the 160 study participants who opted to participate in the add-on smartphone-based study. Of those 160, 38 did not enroll within the mobile app, another 19 enrolled but did not complete a full test session, and 3 were excluded due to no valid test sessions following quality assurance as described above. As a result, 100 individuals who contributed at least 1 valid test session of the MDT-OS (see Table 1 and flow chart in Figure S1) and 66 participants who contributed at least 1 valid test session of the MDT-OS and the ORR task (see Table S1) were included. Following initial filtering, we thus analyzed data from 49 CU Aβ−, 28 CU Aβ+, 9 MCI Aβ−, and 14 MCI Aβ+ in the MDT-OS and from 31 CU Aβ−, 19 CU Aβ+, 7 MCI Aβ−, and 9 MCI Aβ+ in the ORR-DR.

3.2 | Older participants show high acceptability of smartphone-based assessments

First, we were interested in the acceptability of the remote smartphone-based assessments. A total of 120 participants completed a questionnaire about their phone usage. Out of those, 88% owned a smartphone, 64% reported they would download apps by themselves, and 21% with assistance. Participants reported that they had been using a smartphone for 8.2 years on average (SD: 4.5 years, range: 1–25 years), and spend 1.7 h per day using it (SD: 1.6 h, range: 1–10 h).

As reported above, 160 individuals were recruited into the add-on smartphone study. 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

FIGURE 3 Acceptability and user experience. Results from telephone-based interviews with 38 participants focusing on their overall experience with the remote and unsupervised study.

with Internet connection (7%). Seventy-three percent 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 medium 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 to 4.1 and in subjective task difficulty from 2.9 to 3.5). However, as expected, the second phase of the task (i.e., 2-back or delayed retrieval compared to 1-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 and 24.6 cm (mean 13.2 cm, SD = 3.4). While 91% of participants completed the test sessions on a smartphone, 9% used a tablet. Regarding operating systems, 59% used iOS devices, 35% used Android devices, and for 6% the operating system was unknown.

We approached 38 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. Seventy-six percent said they would prefer the mobile tests over in-person paper-and-pencil tests. Seventy-nine percent did not experience their device's screen as too small to see all the details

in the tasks. Eighty-four percent found the number of tests and their duration just right, and 86% rated their experience using the app 7 or higher on a 10-point scale. Out of the 38 participants who completed the phone interview, 36 had finished the study and only 2 participants had dropped out of the study before the end. While those 2 similarly thought that the app was easy to use and the instructions were clear, 1 thought that the screen was too small and both did not enjoy completing the tests in the app (disagree and neutral). One of them found the amount and number of times was just right, while the other rated them as too long and too often (see Figure S2).

Finally, we compared individuals that were not interested in participating, those that agreed to participate but never enrolled in the mobile app, those that enrolled but never completed a valid test session and those that contributed valid test sessions regarding key characteristics to further understand the participating sample. We found that those participants who provided valid test sessions were on average younger than those from all other groups ($F = 8.6, p < 0.001$; $M_{NotInterested} = 72.5$ years (SD 6.38), $M_{NotEnrolled} = 72$ years (SD = 7.36), $M_{NoTests} = 69.2$ years (SD = 8.4), $M_{ValidTests} = 64.5$ (SD = 10.4)). Further, 67% of those that were not interested in participating were MCI patients, while only 26% of individuals across all other groups had an MCI diagnosis. There was no difference in CSF Aβ42/40 levels across groups.

We found no differences between the groups regarding their daily use of smartphones ($F = 1.43, p = 0.239$) and the years they have been using smartphones ($F = 0.44, p = 0.724$). However, 62.5% of those that were not interested in participating or never enrolled in the app had either never installed an app before or had done so only with help. In contrast, 69% of those who enrolled in the app and/or provided valid data had already installed apps on their own.

3.3 | 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. To understand the relationship with demographics and potential confounds, we first calculated multiple regression models with age, sex, years of education, device type (smartphone or tablet), and operating system to identify the associations with both memory paradigms. In addition, we included time-to-retrieval, that is, the time between memory encoding and retrieval, as a covariate for the ORR-DR. For MDT-OS, age and sex were significant predictors for the object, and age and device type were a significant predictor for the scene condition where higher age, female sex, and tablet usage were associated with worse task performance (MDT-O: $\beta_{age} = -0.008$, $p < 0.001$; $\beta_{sex} = -0.08$, $p = 0.045$; $\beta_{education} = 0.006$, $p = 0.287$; $\beta_{device\ type} = -0.13$, $p = 0.0524$; $\beta_{OS} = -0.05$, $p = 0.209$; MDT-S: $\beta_{age} = -0.009$, $p < 0.001$; $\beta_{sex} = -0.08$, $p = 0.062$; $\beta_{education} = 0.008$, $p = 0.230$; $\beta_{device\ type} = -0.18$, $p = 0.015$; $\beta_{OS} = -0.05$, $p = 0.240$). For the ORR-DR, only age was associated with task performance ($\beta_{age} = -0.22$, $p < 0.001$; $\beta_{sex} = -0.24$, $p = 0.814$; $\beta_{education} = 0.08$, $p = 0.617$; $\beta_{delay} = -0.63$, $p = 0.358$; $\beta_{device\ type} = -1.59$, $p = 0.427$; $\beta_{OS} = -1.88$, $p = 0.065$).

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 Functional Magnetic Resonance Imaging scan. While this was a similar task (MDT-OS), 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 using a smartphone. Pearson correlation coefficients revealed a strong relationship between the MDT-OS corrected hit rate across both settings ($r = 0.66$, $p < 0.001$; MDT-O: $r = 0.52$, $p < 0.001$; MDT-S: $r = 0.48$, $p < 0.001$). 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 SDMT, verbal fluency and the MMSE. Due to ceiling effects of ADAS delayed recall and the MMSE in early AD, we assessed Spearman rank correlation rho. All measures were significantly associated with the MDT-OS corrected hit rates (*ADASdelayed*: $\rho = 0.53$, $p < 0.001$; MDT-O: $\rho = 0.54$, $p < 0.001$; MDT-S: $\rho = 0.44$, $p < 0.001$; *MMSE*: $\rho = 0.39$, $p < 0.001$; MDT-O: $\rho = 0.37$, $p < 0.001$; MDT-S: $r = 0.33$, $p < 0.001$; *VerbalFluency*: $r = 0.59$, $p < 0.001$; MDT-O: $r = 0.55$, $p < 0.001$; MDT-S: $r = 0.52$, $p < 0.001$; *SDMT*: $r = 0.58$, $p < 0.001$; MDT-O: $r = 0.54$, $p < 0.001$; MDT-S: $r = 0.50$, $p < 0.001$). Similarly, ORR-DR performance was strongly associated with ADAS delayed recall and also associated with all other outcomes (*ADASdelayed*: $\rho = 0.63$, $p < 0.001$; *MMSE*: $\rho = 0.36$, $p < 0.001$; *VerbalFluency*: $r = 0.66$, $p < 0.001$; *SDMT*: $r = 0.63$, $p < 0.001$).

The mPACC score was associated with MDT-OS (mPACC: $r = 0.62$, $p < 0.001$; MDT-O: $r = 0.58$, $p < 0.001$; MDT-S: $r = 0.54$, $p < 0.001$) and ORR delayed retrieval performance ($r = 0.7$, $p < 0.001$). All relationships above survived corrections for multiple comparisons using False Discovery Rate (FDR) correction.

3.4 | Digital memory scores are associated with measures of AD pathology and MTL 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 MTL-dependent memory processes.[3,11,13] 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,[4,43,44] 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 SUVr in the precuneus representing early $A\beta$ deposition.[45]

Regarding tau-PET, MDT-O, but not MDT-S, was significantly associated with area 35 tau-PET SUVr (MDT-O: $\beta = -0.09$, SE = 0.03, $p = 0.007$; MDT-S: $\beta = -0.04$, SE = 0.04, $p = 0.229$), anterior hippocampal tau-PET SUVr (MDT-O: $\beta = -0.24$, SE = 0.07, $p < 0.001$; MDT-S: $\beta = -0.019$, SE = 0.08, $p = 0.813$), and posterior hippocampal tau-PET SUVr when accounting for participant age and sex (MDT-O: $\beta = -0.17$, SE = 0.08, $p = 0.03$; MDT-S: $\beta = -0.001$, SE = 0.09, $p = 0.989$). Likewise, ORR-DR performance was significantly associated with area 35 tau-PET SUVr ($\beta = -2.3$, SE = 0.7, $p = 0.002$) and anterior ($\beta = -4.44$, SE = 1.53, $p = 0.005$) but not posterior hippocampal tau-PET SUVr ($\beta = -2.99$, SE = 1.86, $p = 0.113$) when accounting for participant age, sex and time between encoding and retrieval. These relationships survived corrections for multiple comparisons using FDR correction. However, when additionally accounting for $A\beta$-PET SUVr in precuneus, only the relationship between the MDT-O and the anterior hippocampus remained statistically significant (MDT-O: $\beta = -0.22$, SE = 0.09, $p = 0.013$) while the relationship of the MDT-O with area 35 and posterior hippocampal tau-PET weakened (area 35: $\beta = -0.07$, SE = 0.04, $p = 0.113$; posterior hippocampus: $\beta = -0.11$, SE = 0.1, $p = 0.242$) as did the relationship of the ORR-DR and area 35 and anterior hippocampal tau-PET SUVr (area 35: $\beta = -1.54$, SE = 0.97, $p = 0.117$; anterior hippocampus: $\beta = -2.37$, SE = 2.02, $p = 0.245$)

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: $\beta = -0.18$, SE = 0.08, $p = 0.026$; MDT-S: $\beta = -0.08$, SE = 0.09, $p = 0.343$; ORR: $\beta = -5.66$, SE = 01.85, $p = 0.003$). However, the $A\beta$-PET SUVr relationships did not hold when accounting for area 35 tau-PET SUVR (all p-values > 0.26).
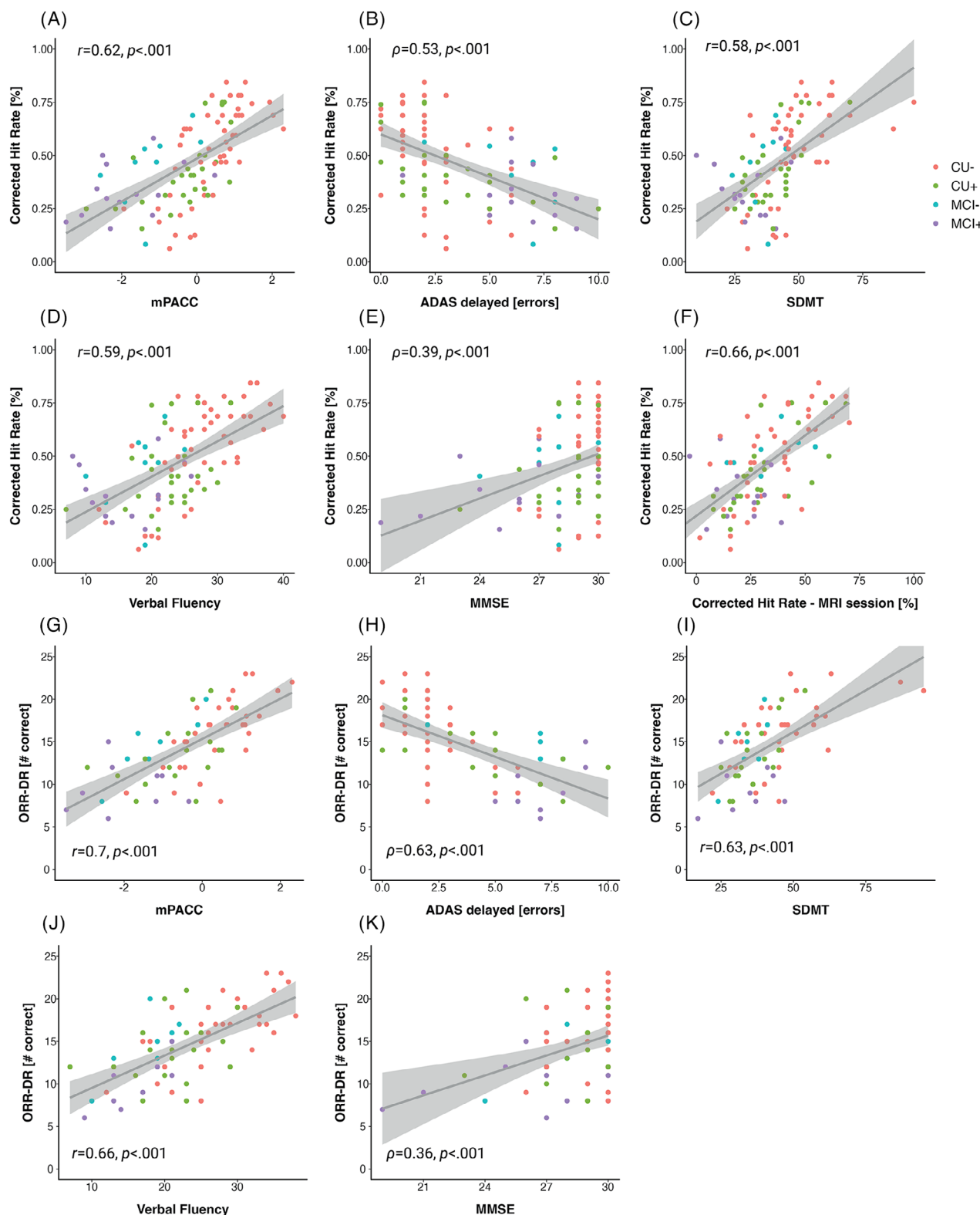
FIGURE 4 Construct validity of the MDT-OS and the ORR. Scatter plots showing relationships of the MDT-OS Corrected hit rate with the (A) modified Preclinical Alzheimer's Cognitive Composite, (B) errors in the delayed 10-word list recall test from ADAS-cog, (C) Symbol Digit Modalities Test, (D) animal 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-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. ADAS, Alzheimer's Disease Assessment Scale; ADAS-cog, Alzheimer's Disease Assessment Scale—Cognitive Subscale; MDT-OS; mnemonic discrimination test for objects and scenes; ORR, Objects-In-Room Recall; ORR-DR, ORR delayed recall score.
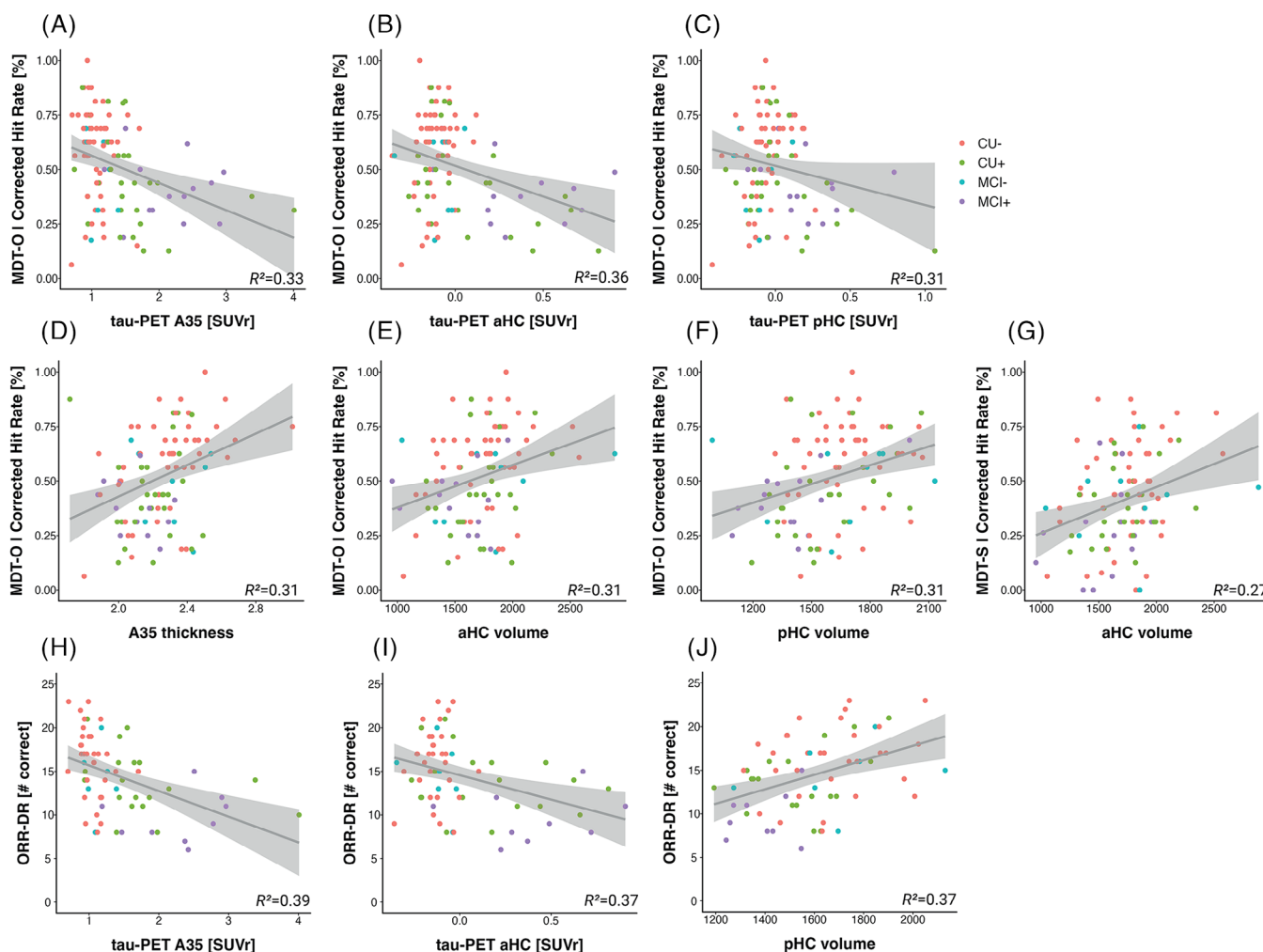
FIGURE 5 Criterion validity of the MDT-OS and the ORR-DR. Scatter plots showing relationships between the MDT-O and tau-PET uptake in (A) area 35, (B) the anterior and (C) posterior hippocampus, as well as cortical thickness in area 35 (D) and anterior (E) and posterior hippocampal volume (F). Panel G shows the relationship of the MDT-S with anterior hippocampal volume. Finally, panel (H), (I), and (J) 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. 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; PET, positron emission tomography.

With respect to MTL measures of atrophy, MDT-O was significantly associated with area 35 thickness as well as anterior and posterior hippocampal volume (area 35: $\beta = 0.22$, SE = 0.11, $p = 0.044$ [robust regression: $F = 5.34$, $p = 0.022$]; anterior hippocampus: $\beta = 301.3$, SE = 150.8, $p = 0.049$; posterior hippocampus: $\beta = 218$, SE = 108, $p = 0.046$), while MDT-S was only associated with anterior hippocampal volume (anterior hippocampus: $\beta = 420.4$, SE = 134.4, $p = 0.002$, A35: $\beta = 0.17$, SE = 0.1, $p = 0.081$, posterior hippocampus: $\beta = 190.7$, SE = 99.2, $p = 0.058$) when accounting for age, sex, and intracranial volume. ORR-DR performance was significantly associated with posterior hippocampal volume ($\beta = 15$, SE = 6, $p = 0.015$) but neither anterior hippocampal volume ($\beta = 15.6$, SE = 8.03, $p = 0.057$) nor area 35 thickness ($\beta = 0.001$, SE = 0.006, $p = 0.076$) 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. Taken together, both memory tasks showed sensitivity to measures of AD pathology as well as MTL atrophy as a measure of neurodegeneration.

3.5 | 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 cognitive decline. Recent work has shown that plasma p-tau217 was associated with cognitive decline as well as disease progression.[46,47] 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 to test which model best predicted 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 Tables 2 and 3, see also Table S2 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 (β [SE] = −0.024 [0.005]; $p < 0.001$), MDT-S (β [SE] = 0.174 [0.032]; $p < 0.001$), age (with higher age associated with worse mPACC slopes; β [SE] = −0.003 [0.001]; $p = < 0.001$), sex (with male sex associated with worse mPACC slopes; β [SE] = 0.044 [0.013]; $p < 0.001$), and education (β [SE] = −0.004 [0.002]; $p = 0.069$; AIC for the overall model: −474; R^2 = 0.31, 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 Table S1 for sample characteristics), the best combination model for prediction of mPACC slopes included plasma p-tau217 (β [SE] = −0.03 [0.006]; $p < 0.001$), ORR-DR (β [SE] = 0.012 [0.003]; $p < 0.001$), MDT-S (β [SE] = 0.167 [0.041]; $p < 0.001$), age (β [SE] = 0.000 [0.001]; $p = 0.858$), sex (with male sex associated with worse mPACC slopes; β [SE] = 0.048 [0.016]; $p = 0.004$), and education (with higher education associated with worse mPACC slopes; (β [SE] = −0.009 [0.003]; $p < 0.001$; AIC for the overall model: −298; R^2 = 0.42, 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.

3.6 | 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 who had at least 4 sessions completed ($n = 73$ and $n = 37$ completed repeated assessments across 4 sessions of the MDT-OS and ORR respectively) and calculated 2-way random ICC for 2 scenarios. Regarding a scenario where single test sessions would be used, we calculated the ICC between the first and second session of a respective test and regarding a scenario where the average between 2 sessions would be used, we calculated the ICC between the average of the first and second session of a respective test and the average the third and fourth session of a respective test. A single session of MDT-OS (ICC of 0.65, 95% CI [0.5, 0.77]; MDT-O: 0.58 [CI 0.4, 0.71]; MDT-S: 0.52 [CI 0.33, 0.67]; $n = 73$) and a single session of ORR (ICC of 0.67 [CI 0.41, 0.8]; $n = 37$) both showed moderate retest reliability. In the second scenario, the average of 2 test sessions of MDT-OS (ICC of 0.83 [CI 0.7, 0.89]; MDT-O: 0.65 [CI 0.49, 0.76];

MDT-S: 0.68 [CI 0.53, 0.78]), as well as the average of two test sessions of ORR (ICC of 0.79 [CI 0.63, 0.89]) showed good retest reliability.

4 | DISCUSSION

We found good construct validity of remote and unsupervised digital memory assessments when comparing them with in-clinic traditional neuropsychological assessments, and that baseline performance in the remote memory assessments was associated with decline in the mPACC score. The model that best predicted cognitive decline in the mPACC included a combination of plasma p-tau217 and MDT-S. In a sub-sample of individuals where both digital cognitive 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 tau-PET as well as downstream MRI measures of MTL atrophy. Finally, the onboarding into the study as well as the unsupervised completion of memory tasks itself was rated positively by participants.

4.1 | 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 unsupervised studies using digital devices. However, recent work has shown that, while older adults were indeed less familiar with technology, many older participants decided to participate in remote studies and showed exceptional adherence when studies were planned thoughtful and included user-centered design.[48] In line with this, we found that a large majority of older participants owned a mobile device and downloaded apps by themselves or with help. However, we found that MCI and older age were more common in individuals who were not interested in participating or did not contribute complete datasets. Many of those who did not participate indicated that they had either never installed an app before or only with help, while the majority of participants who contributed complete datasets had already installed apps on their own. While only 10% were not interested in participating in the remote assessments, more than 20% of individuals who initially agreed to participate did never enroll within the app. Therefore, there seems to be a large potential to encourage participation by offering support to those with MCI, that are older and those with less experience with smartphones as has been demonstrated in earlier work.[48]

Participants rated both remote memory paradigms as challenging but not too difficult. A large majority found the instructions clear and the application easy to use, and the app was overall rated at least 7 out of 10 by more than 85% of participants. This indicates that the app was easy to work with for most participants. Importantly, however, these interviews almost exclusively included participants who had

TABLE 2 Results of multiple regression models in the entire sample ($n = 86$[a]) predicting future cognitive decline including baseline plasma p-tau217 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-Value	Estimate	CI	p-Value	Estimate	CI	p-Value	Estimate	CI	p-Value
(intercept)	0.317	0.203–0.432	**<0.001**	0.149	0.015–0.284	**0.029**	0.316	0.205 to 0.426	**<0.001**	0.179	0.062 to 0.296	**0.003**
Age (years)	−0.005	−0.006 to −0.004	**<0.001**	−0.003	−0.005 to −0.002	**<0.001**	−0.004	−0.005 to −0.003	**<0.001**	−0.003	−0.004 to −0.001	**<0.001**
Sex (female)	0.031	0.004 to 0.059	**0.023**	0.050	0.023 to 0.077	**<0.001**	0.029	0.002 to 0.055	**0.032**	0.044	0.019 to 0.070	**0.001**
Education (years)	0.000	−0.004 to 0.004	0.975	−0.003	−0.007 to 0.001	0.192	−0.001	−0.005 to 0.003	0.649	−0.004	−0.008 to 0.000	0.069
MDT-O				0.048	−0.034 to 0.130	0.251						
MDT-S				0.149	0.073 to 0.225	**<0.001**				0.174	0.111 to 0.237	**<0.001**
p-tau217							−0.024	−0.034 to −0.014	**<0.001**	−0.024	−0.034 to −0.014	**<0.001**
Observations	292			292			292			292		
R^2/R^2 adj	0.180/0.172			0.252/0.239			0.235/0.225			0.306/0.294		
AIC	−429.038			−451.968			−447.381			−473.811		

Abbreviations: AIC, Akaike information criterion; BBM, blood-based biomarker; CI, confidence interval; MDT-O, Mnemonic Discrimination Task for Objects; MDT-S, Mnemonic Discrimination Task for Scenes; mPACC, modified Preclinical Alzheimer's Cognitive Composite.

Bold values significance $p < 0.05$.

[a] Note that 14 participants did not have p-tau217 levels available.

TABLE 3 Results of multiple regression models in a restricted sample ($n = 55$[a]) where both digital markers were completed predicting future cognitive decline including baseline plasma p-tau217 and digital remote cognitive 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-Value	Estimate	CI	p-Value	Estimate	CI	p-Value	Estimate	CI	p-Value
(intercept)	0.328	0.175 to 0.482	**<0.001**	−0.133	−0.315 to 0.049	0.151	0.314	0.166 to 0.461	**<0.001**	−0.106	−0.275 to 0.064	0.220
Age (years)	−0.005	−0.007 to −0.003	**<0.001**	−0.001	−0.003 to 0.001	0.475	−0.004	−0.005 to −0.002	**<0.001**	0.000	−0.002 to 0.002	0.858
Sex (female)	0.043	0.005 to 0.080	**0.027**	0.059	0.025 to 0.094	**0.001**	0.033	−0.003 to 0.070	0.076	0.048	0.015 to 0.080	**0.004**
Education (years)	−0.002	−0.007 to 0.004	0.582	−0.008	−0.013 to −0.003	**0.004**	−0.003	−0.008 to 0.003	0.338	−0.009	−0.014 to −0.004	**<0.001**
MDT-O				0.089	−0.021 to 0.198	0.113						
MDT-S				0.102	0.009 to 0.195	**0.033**				0.167	0.086 to 0.249	**<0.001**
ORR-DR				0.013	0.007 to 0.018	**<0.001**				0.012	0.007 to 0.017	**<0.001**
p-tau217							−0.028	−0.041 to −0.014	**<0.001**	−0.028	−0.041 to −0.016	**<0.001**
Observations	187			187			187			187		
R^2/R^2 adj	0.151/0.138			0.358/0.337			0.219/0.201			0.417/0.398		
AIC	−233.573			−279.850			−246.965			−297.803		

Abbreviations: AIC, Akaike information criterion; BBM, blood-based biomarker; CI, confidence interval; MDT-O, Mnemonic Discrimination Task for Objects; MDT-S, Mnemonic Discrimination Task for Scenes; mPACC, modified Preclinical Alzheimer's Cognitive Composite; ORR-DR, Object-In-Room Recall–Delayed Recall.

Bold values significance $p < 0.05$.

[a]Note that 11 participants did not have p-tau217 levels available.

finished the study after 1 year which might have led to biased results. Future studies thus need to incorporate usability interviews earlier in the study schedule.

4.2 | Remote and unsupervised assessments reflect traditional neuropsychological assessments

When introducing a novel cognitive measure, 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 unstandardized environments of the participant's choice. Recent work using smartphone-based assessments in samples of older participants could demonstrate high construct validity across various cognitive domains.[20–22,49,50] Both the ORR and the MDT-OS measure visual memory, which makes it difficult to cheat, as would be possible in verbal memory tasks, where participants could, for example, take notes. While the ORR specifically aims to measure delayed memory, the MDT-OS measures precision memory in an n-back task design known to also rely on attentional and executive processes.[51] We found that the ORR-DR and MDT-OS were strongly associated with ADAS delayed recall, SDMT, and verbal fluency and less so with the MMSE. The strongest relationship for both measures was found with the mPACC. These findings support earlier work using the neotiv memory tasks in samples with CU and MCI patients.[21,28] 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 cognitive decline.

While high construct validity indicates that the tasks were successfully designed to measure memory function, it also indicates that study participants were successful in completing the 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 sessions were reported without any distractions. Interestingly, our findings show that CU individuals reported more distractions compared to MCI patients, in line with earlier research.[52]

4.3 | Remote and unsupervised assessments are sensitive to measures of AD pathology

Both the MDT-OS and the ORR are considered MTL-dependent tasks[3,11,13,15,16] and earlier work has already linked performance in the MDT-OS with fluid and imaging measures of AD pathology.[3,13] While the MDT-O is associated with an object pathway including the perirhinal cortex (including area 35) and the anterior-lateral entorhinal cortex, the MDT-S targets a scene pathway including the parahippocampal cortex and the posterior-medial entorhinal cortex.[3,11,13] Elevated tau-PET signal in early disease stages can be found in perirhinal cortex and the anterior-lateral entorhinal cortex before it progresses toward

the hippocampus.[4] Thus, we would expect relationships and effects of tau pathology specifically on object memory as we have seen in earlier work.[3,13] 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, and there was no such relationship for the MDT-S.[13] However, these effects weakened, and several became non-significant when we added a measure of amyloid burden to the models. While this is in part expected given the high correlation between tau and amyloid burden, future analyses with bigger sample sizes need to confirm these relationships. 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.[4,53,54] 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 that task performance is affected by early accumulation of AD pathology.

4.4 | Potential of remote assessments for case finding, monitoring, and prognosis of cognitive impairment in AD

Remote unsupervised digital cognitive assessments hold promise for case finding in healthcare 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[28] and potentially even Aβ-positive CU participants.[19,26] Regarding prognosis, recent work has shown that plasma biomarkers for tau pathology are associated with future cognitive decline,[46] and in combination with brief in-clinic pen-and-paper cognitive tests, can identify MCI patients who are likely to progress toward dementia.[47] Similarly, Tsoy and colleagues recently showed that a combination of plasma biomarkers and brief in-clinic digital cognitive assessments could predict Aβ-positivity and were associated with concurrent disease severity as well as future functional decline.[55] 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 memory assessments can significantly contribute to the prediction of future decline. Moreover, the moderate-to-good retest reliability of both memory assessments, coupled with the minimal practice effects due to the utilization of parallel test sets[24] 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 need to 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. First, the modest sample size, particularly for participants who completed both memory paradigms. Second, 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.

5 | 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.

CONFLICT OF INTEREST STATEMENT

O.H. has acquired research support (for the institution) from ADx, AVID Radiopharmaceuticals, Biogen, Eli Lilly, Eisai, Fujirebio, GE Healthcare, Pfizer, and Roche. In the past 2 years, he has received consultancy/speaker fees from AC Immune, Amylyx, Alzpath, BioArctic, Biogen, Cerveau, Eisai, Eli Lilly, Fujirebio, Merck, Novartis, Novo Nordisk, Roche, Sanofi and Siemens. S.P. 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. E.D. reports personal fees from Biogen, Roche, Lilly, Eisai and UCL Consultancy as well as non-financial support from Rox Health. D.B. and E.D. are scientific co-founders of neotiv GmbH and own company shares. The other authors report no competing interests. Author disclosures are available in the supporting information.

DATA AVAILABILITY STATEMENT

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 if 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.

CONSENT STATEMENT

All human participants provided informed consent to participate.

REFERENCES

1. Hansson O. Biomarkers for neurodegenerative diseases. *Nat Med*. 2021;27(6):954-963. doi:10.1038/s41591-021-01382-x
2. Hansson O, Blennow K, Zetterberg H, Dage J. Blood biomarkers for Alzheimer's disease in clinical practice and trials. *Nat Aging*. 2023;3(5):506-519. doi:10.1038/s43587-023-00403-3
3. Maass A, Berron D, Harrison TM, et al. Alzheimer's pathology targets distinct memory networks in the ageing brain. *Brain*. 2019;142(8):2492-2509. doi:10.1093/brain/awz154
4. Berron D, Vogel JW, Insel PS, et al. Early stages of tau pathology and its associations with functional connectivity, atrophy and memory. *Brain*. 2021;144(9):awab114. doi:10.1093/brain/awab114
5. Grothe MJ, Barthel H, Sepulcre J, et al. In vivo staging of regional amyloid deposition. *Neurology*. 2017;89(20):2031-2038. doi:10.1212/wnl.0000000000004643
6. Ritchey M, Cooper RA. Deconstructing the posterior medial episodic network. *Trends Cogn Sci*. 2020;24(6):451-465. doi:10.1016/j.tics.2020.03.006
7. Ranganath C, Ritchey M. Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci*. 2012;13(10):713-726. doi:10.1038/nrn3338
8. Berron D, Schütze H, Maass A, et al. Strong evidence for pattern separation in human dentate gyrus. *J Neurosci*. 2016;36(29):7569-7579. doi:10.1523/jneurosci.0518-16.2016
9. Bakker A, Kirwan BC, Miller M, Stark CE. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*. 2008;319(5870):1640-1642. doi:10.1126/science.1152882
10. Grande X, Berron D, Horner AJ, Bisby JA, Düzel E, Burgess N. Holistic recollection via pattern completion involves hippocampal subfield CA3. *J Neurosci*. 2019;39(41):8100-8111. doi:10.1523/jneurosci.0722-19.2019
11. Berron D, Neumann K, Maass A, et al. Age-related functional changes in domain-specific medial temporal lobe pathways. *Neurobiol Aging*. 2018;65:86-97. doi:10.1016/j.neurobiolaging.2017.12.030
12. Maass A, Berron D, Libby LA, Ranganath C, Düzel E. Functional subregions of the human entorhinal cortex. *eLife*. 2015;4:e06426. doi:10.7554/elife.06426
13. Berron D, Cardenas-Blanco A, Bittner D, et al. Higher CSF tau levels are related to hippocampal hyperactivity and object mnemonic discrimination in older adults. *J Neurosci*. 2019;39(44):8788-8797. doi:10.1523/jneurosci.1279-19.2019

14. Schröder T, Haak KV, Jimenez NI, Beckmann CF, Doeller CF. Functional topography of the human entorhinal cortex. *eLife.* 2015;4:e06738. doi:10.7554/elife.06738

15. Grande X, Sauvage MM, Becke A, Düzel E, Berron D. Transversal functional connectivity and scene-specific processing in the human entorhinal-hippocampal circuitry. *eLife.* 2022;11:e76479. doi:10.7554/elife.76479

16. Grande X, Berron D, Maass A, Bainbridge W, Düzel E. Content-specific vulnerability of recent episodic memories in Alzheimer's disease. *Neuropsychologia.* 2021;160:107976. doi:10.1016/j.neuropsychologia.2021.107976

17. Masanneck L, Gieseler P, Gordon WJ. et al. Evidence from Clinical-Trials.gov on the growth of Digital Health Technologies in neurology trials. *npj Digit. Med.* 2023;6:23. doi: 10.1038/s41746-023-00767-1

18. Öhman F, Hassenstab J, Berron D, Schöll M, Papp KV. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimers Dement Diagn Assess Dis Monit.* 2021;13(1):e12217. doi:10.1002/dad2.12217

19. Samaroo A, Amariglio RE, Burnham S, et al. Diminished Learning Over Repeated Exposures (LORE) in preclinical Alzheimer's disease. *Alzheimers Dement Diagn Assess Dis Monit.* 2020;12(1):e12132. doi:10.1002/dad2.12132

20. Papp KV, Samaroo A, Chou H, et al. Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimers Dement Diagn Assess Dis Monit.* 2021;13(1):e12243. doi:10.1002/dad2.12243

21. Öhman F, Berron D, Papp KV, et al. Unsupervised mobile app-based cognitive testing in a population-based study of older adults born 1944. *Frontiers Digital Heal.* 2022;4:933265. doi:10.3389/fdgth.2022.933265

22. Nicosia J, Aschenbrenner AJ, Balota DA, et al. 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.* 2022;29(5):459-471. doi:10.1017/s135561772200042x

23. Thompson LI, Harrington KD, Roque N, et al. A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults. *Alzheimers Dement Diagn Assess Dis Monit.* 2022;14(1):e12283. doi:10.1002/dad2.12283

24. Berron D, Ziegler G, Vieweg P, et al. Feasibility of digital memory assessments in an unsupervised and remote study setting. *Front Digit Heal.* 2022;4:892997. doi:10.3389/fdgth.2022.892997

25. Berron D, Glanz W, Clark L, et al. A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. *NPJ Digit Med.* 2024;7(1):79. doi:10.1038/s41746-024-00999-9

26. Jutten RJ, Rentz DM, Fu JF, et al. Monthly at-home computerized cognitive testing to detect diminished practice effects in preclinical Alzheimer's disease. *Front Aging Neurosci.* 2022;13:800126. doi:10.3389/fnagi.2021.800126

27. Thompson LI, Kunicki ZJ, Emrani S, et al. Remote and in-clinic digital cognitive screening tools outperform the MoCA to distinguish cerebral amyloid status among cognitively healthy older adults. *Alzheimer's Dement..* 2023;15:e12500. 10.1002/dad2.12500

28. Berron D, Glanz W, Clark L, et al. A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. *NPJ Digit Med.* 2024;7:79. doi:10.1038/s41746-024-00999-9

29. Quadalti C, Palmqvist S, Hall S, et al. Clinical effects of Lewy body pathology in cognitively impaired individuals. *Nat Med.* 2023;29(8):1964-1970. doi:10.1038/s41591-023-02449-7

30. Palmqvist S, Janelidze S, Quiroz YT, et al. Discriminative Accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *JAMA.* 2020;324(8):772-781. doi:10.1001/jama.2020.12134

31. Cho SH, Choe YS, Park S, et al. Appropriate reference region selection of 18F-florbetaben and 18F-flutemetamol beta-amyloid PET expressed in centiloid. *Sci Rep.* 2020;10(1):14950. doi:10.1038/s41598-020-70978-z

32. Xie L, Pluta JB, Das SR, et al. Multi-template analysis of human perirhinal cortex in brain MRI: explicitly accounting for anatomical variability. *NeuroImage.* 2017;144:183-202.

33. Baker SL, Maass A, Jagust WJ. Considerations and code for partial volume correcting [18F]-AV-1451 tau PET data. *Data Brief.* 2017;15(Acta Neuropathol. 82 1991):648-657. doi:10.1016/j.dib.2017.10.024

34. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. *J Nucl Med.* 1998;39(5):904-911.

35. Thomas BA, Erlandsson K, Modat M, et al. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur J Nucl Med Mol I.* 2011;38(6):1104-1119. doi:10.1007/s00259-011-1745-9

36. Thurfjell L, Lilja J, Lundqvist R, et al. 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.* 2014;55(10):1623-1628. doi:10.2967/jnumed.114.142109

37. Mack WJ, Freed DM, Williams BW, Henderson VW. Boston Naming Test: shortened versions for use in Alzheimer's disease. *J Gerontol.* 1992;47(3):P154-P158. doi:10.1093/geronj/47.3.p154

38. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state" a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12(3):189-198. doi:10.1016/0022-3956(75)90026-6

39. Wechsler D. *WAIS-R Manual: Wechsler Adult Intelligence Scale-Revised.* Psychological Corporation; 1981.

40. Papp KV, Rofael H, Veroff AE, et al. Sensitivity of the Preclinical Alzheimer's Cognitive Composite (PACC), PACC5, and Repeatable Battery for Neuropsychological Status (RBANS) to amyloid status in preclinical Alzheimer's disease -atabecestat phase 2b/3 EARLY clinical trial. *J Prev Alzheimer's Dis.* 2022;9(2):255-261. doi:10.14283/jpad.2022.17

41. Burnham KP, Anderson DR. Multimodel inference. *Sociol Methods Res.* 2004;33(2):261-304. doi:10.1177/0049124104268644

42. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

43. de Flores R, Das SR, Xie L, et al. Medial temporal lobe networks in Alzheimer's disease: structural and molecular vulnerabilities. *J Neurosci.* 2022;42(10):2131-2141. doi:10.1523/jneurosci.0949-21.2021

44. Sanchez JS, Becker JA, Jacobs HIL, et al. The cortical origin and initial spread of medial temporal tauopathy in Alzheimer's disease assessed with positron emission tomography. *Sci Transl Med.* 2021;13(577):eabc0655. doi:10.1126/scitranslmed.abc0655

45. Palmqvist S, Schöll M, Strandberg O, et al. Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nat Commun.* 2017;8(1):1214. doi:10.1038/s41467-017-01150-x

46. Mattsson-Carlgren N, Salvadó G, Ashton NJ, et al. Prediction of longitudinal cognitive decline in preclinical Alzheimer disease using plasma biomarkers. *JAMA Neurol.* 2023;80(4):360-369. doi:10.1001/jamaneurol.2022.5272

47. Palmqvist S, Tideman P, Cullen N, et al. Prediction of future Alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. *Nat Med.* 2021;27(6):1034-1042. doi:10.1038/s41591-021-01348-z

48. Nicosia J, Aschenbrenner AJ, Adams SL, et al. Bridging the technological divide: stigmas and challenges with technology in digital brain health studies of older adults. *Frontiers Digital Heal.* 2022;4:880055. doi:10.3389/fdgth.2022.880055

49. Singh S, Strong R, Xu I, et al. Ecological momentary assessment of cognition in clinical and community samples: reliability and validity study. *J Med Internet Res.* 2023;25:e45028. doi:10.2196/45028

50. Vyshedskiy A, Netson R, Fridberg E, et al. Boston cognitive assessment (BOCA) — a comprehensive self-administered smartphone- and computer-based at-home test for longitudinal tracking of cognitive performance. *BMC Neurol.* 2022;22(1):92. doi:10.1186/s12883-022-02620-6

51. Gajewski PD, Hanisch E, Falkenstein M, Thönes S, Wascher E. What does the n-back task measure as we get older? *Front Psychol.* 2018;9:2208. doi:10.3389/fpsyg.2018.02208

52. Madero EN, Anderson J, Bott NT, et al. Environmental Distractions during unsupervised remote digital cognitive assessment. *J Prev Alzheimers Dis.* 2021;8(3):263-266. doi:10.14283/jpad.2021.9

53. Xie L, Das SR, Wisse L, et al. Early tau burden correlates with higher rate of atrophy in transentorhinal cortex. *J Alzheimers dis.* 2018;62(1):85-92. doi:10.3233/jad-170945

54. Xie L, Wisse LEM, Das SR, et al. Longitudinal atrophy in early Braak regions in preclinical Alzheimer's disease. *Hum Brain Mapp.* 2020;41(16):4704-4717. doi:10.1002/hbm.25151

55. Tsoy E, Joie RL, VandeVrede L, et al. Scalable plasma and digital cognitive markers for diagnosis and prognosis of Alzheimer's disease and related dementias. *Alzheimers Dement.* 2024;20:2089-2101. doi:10.1002/alz.13686

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Berron D, Olsson E, Andersson F, et al. Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease. *Alzheimer's Dement.* 2024;20:4775–4791. https://doi.org/10.1002/alz.13919