

<https://doi.org/10.1038/s41746-025-01583-5>

A scoping review of remote and unsupervised digital cognitive assessments in preclinical Alzheimer's disease



Sarah E. Polk¹✉, Fredrik Öhman^{2,3,4}, Jason Hassenstab^{5,6}, Alexandra König^{7,8,9}, Kathryn V. Papp¹⁰, Michael Schöll^{2,3,4,11,14} & David Berron^{1,12,13,14}✉

Characterizing subtle cognitive changes in preclinical Alzheimer's disease (AD) is difficult using traditional neuropsychological assessments. Remote and unsupervised digital assessments can improve scalability, measurement reliability, and ecological validity, enabling the capture of subtle changes. We evaluate such tools for use in preclinical AD, or cognitively unimpaired individuals with abnormal levels of AD pathology. We screened 1904 reports for studies remotely assessing cognition in preclinical AD samples. Twenty-three tools were identified, and their usability, reliability, and validity, including construct and criterion validity based on in-person neuropsychological and A β /tau measures, was reported. We present a necessary update to a rapidly evolving field, following our previous review (Öhman et al., 2021) and address open questions of feasibility and reliability of remote testing in older adults. Future applications of such tools are discussed, including longitudinal monitoring of cognition, scalable case finding, and individualized prognostics in both clinical trials and healthcare contexts.

Subtle cognitive changes may already emerge in the preclinical stage of Alzheimer's disease (AD), that is, in cognitively unimpaired individuals with abnormal levels of biomarkers indicative of AD pathology, such as amyloid- β (A β) and tau (clinical stage 2 according to ref. 1). Determining whether an individual has abnormal levels of A β and/or tau, and thereby establishing a probable etiology for cognitive decline and later impairment, can now reliably be achieved using fluid and imaging biomarkers^{1,2}. However, the characterization of cognitive changes has proven to be challenging using conventional pen-and-paper neuropsychological assessments^{3–5}. This presents a non-negligible and potentially costly hurdle, as clinical trials aim to

screen and enroll participants at risk for disease-driven cognitive decline before the clinical symptoms of AD appear⁶. Indeed, Langford and colleagues⁷ estimated that using sensitive cognitive measures to pre-screen individuals before ordering PET imaging could have saved over 3.5 M USD in the A4 study. Additionally, although there is yet no medical intervention approved for preclinical AD, the use of anti-amyloid drugs in preclinical samples is already being tested in studies such as AHEAD 3-45 (NCT04468659) and TRAILBLAZER-ALZ 3 (NCT05026866). Once early treatments become available, tools that can reliably detect, predict, and monitor cognitive decline due to AD on an individual level will be essential⁸.

¹Clinical Cognitive Neuroscience, German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany. ²Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden. ³Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden. ⁴Region Västra Götaland, Sahlgrenska University Hospital, Department of Neuropsychiatry, Gothenburg, Sweden. ⁵Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA. ⁶Department of Neurology, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. ⁷ki:elements UG, Saarbrücken, Germany. ⁸Cognition Behaviour Technology (CoBTek) Lab, University Côte d'Azur, Nice, France. ⁹Université Côte d'Azur, Centre Hospitalier et Universitaire, Clinique Gériatrique du Cerveau et du Mouvement, Centre Mémoire de Ressources et de Recherche, Nice, France. ¹⁰Mass General Brigham, Department of Neurology, Harvard Medical School, Boston, MA, USA.

¹¹Dementia Research Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK.

¹²Center for Behavioral Brain Sciences, Otto-von-Guericke University Magdeburg, Magdeburg, Germany. ¹³Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Lund, Sweden. ¹⁴These authors contributed equally: Michael Schöll, David Berron. ✉e-mail: sarah.polk@dzne.de; david.berron@dzne.de

Remote and unsupervised digital cognitive assessments have already shown promise as a widely accessible solution that may be sensitive to subtle cognitive changes related to the initial build-up of AD pathology⁹, even before clinical symptoms manifest (i.e., preclinical AD). So far, in the context of preclinical AD, remote and unsupervised digital cognitive assessments have been exclusively used in research, for example, to characterize cross-sectional cognitive differences between healthy and AD samples (e.g., refs. 10–16). Studies have also thus far focused on establishing the feasibility of remote and unsupervised data collection in samples of older adults with varying degrees of cognitive impairment (e.g., refs. 17–22). Moving forward, remote and unsupervised digital cognitive assessments have great potential in future research directions as well as health-care use cases for sensitive cognitive assessments, including in the identification of individuals with subtle cognitive changes (e.g., as a pre-screening for biomarker testing), the quantification of risk of later cognitive decline in individuals (e.g., prognosis), and the characterization of change in cognition longitudinally using frequent assessments (e.g., to monitor cognitive decline and/or treatment outcomes).

Digital data collection, whether remote or in-person, offers a number of advantages over pen-and-paper testing, such as improved measurement precision (e.g., reaction time, automatic scoring)⁹. A wide range of existing pen-and-paper assessments have been digitized^{23,24}, and a growing number of cognitive tests are being developed specifically for digital administration (e.g., refs. 25–27). In addition to tests that capture conventional cognitive constructs (e.g., episodic memory, processing speed), novel metrics quantifying cognitive function have also been developed, for example, using speech-based tasks (e.g., refs. 28,29) or multi-modal assessments (i.e., assessment of both active and passive markers of cognition simultaneously³⁰). Notably, many tools that use cognitive tests that may be considered conventional can also capture novel metrics (e.g., learning curves using an episodic memory task) due to the digital nature of the tests; thus, this is not a mutually exclusive categorization. Finally, passive data collection can be implemented in participants' natural environments (e.g., using home-monitoring, wearables), increasing the ecological validity of such metrics by allowing the collection of continuous data streams during individuals' everyday lives^{31,32}. See Fig. 1 for an overview of the types of digital cognitive assessments.

The digitization of cognitive assessments enables the remote and unsupervised collection of cognitive data using participants' own devices, which is associated with a number of benefits and challenges. Anyone with a smart device and a network connection can complete remote automated testing. Indeed, data from millions of users across the globe have already been collected using mobile chat- and game-based (e.g., refs. 27,33) as well as web-based cognitive tests (e.g., TestMyBrain). This approach offers many benefits, including cost-effectiveness and reduction of patient burden while still allowing for a high volume of data to be collected, but also many challenges. For example, remote versus in-person data collection is associated with reduced costs related to travel, clinic or laboratory space, and test administration³⁴, though there may be other costs associated with technical support and data servers. Remote data collection can also be less burdensome to patients and participants, as scheduling conflicts (e.g., going into a clinic during business hours) can typically be eliminated, and transportation to and from testing sites is not necessary, though participant burden will also depend on the remote testing schedule.

In addition to efficient data collection from more individuals, the frequency of within-person testing can be increased with remote testing. Participants can feasibly self-administer remote assessments once a month for a year¹⁰, on a daily basis^{18,35}, or even multiple times a day^{36–38}. Measurement burst designs can also be implemented (e.g., a week of daily assessments every six months for several years)^{39,40}. High-frequency testing not only improves the temporal scale on which changes can be detected, but also crucially increases measurement reliability and the sensitivity to intra-individual variability in performance^{41–43}. Notably, many remote and unsupervised paradigms that do not explicitly test learning of specific stimuli use randomized stimulus pairs or parallel versions when repeatedly administering tests to reduce retest effects^{13,20,37,44–47}. Finally, higher frequency within-person testing has opened the door for novel paradigms measuring cognitive processes that are otherwise difficult to capture, such as learning of repeated stimuli^{11,12,35}, recall after several hours or days^{46,48,49}, or the effects of time of day¹³.

Another benefit to unsupervised cognitive testing is an increase in ecological validity. One shortcoming of in-clinic tests is potentially discrepant cognitive performance in clinical settings versus at-home, known as the “white-coat effect”^{50,51}. Performance on remote assessments may be

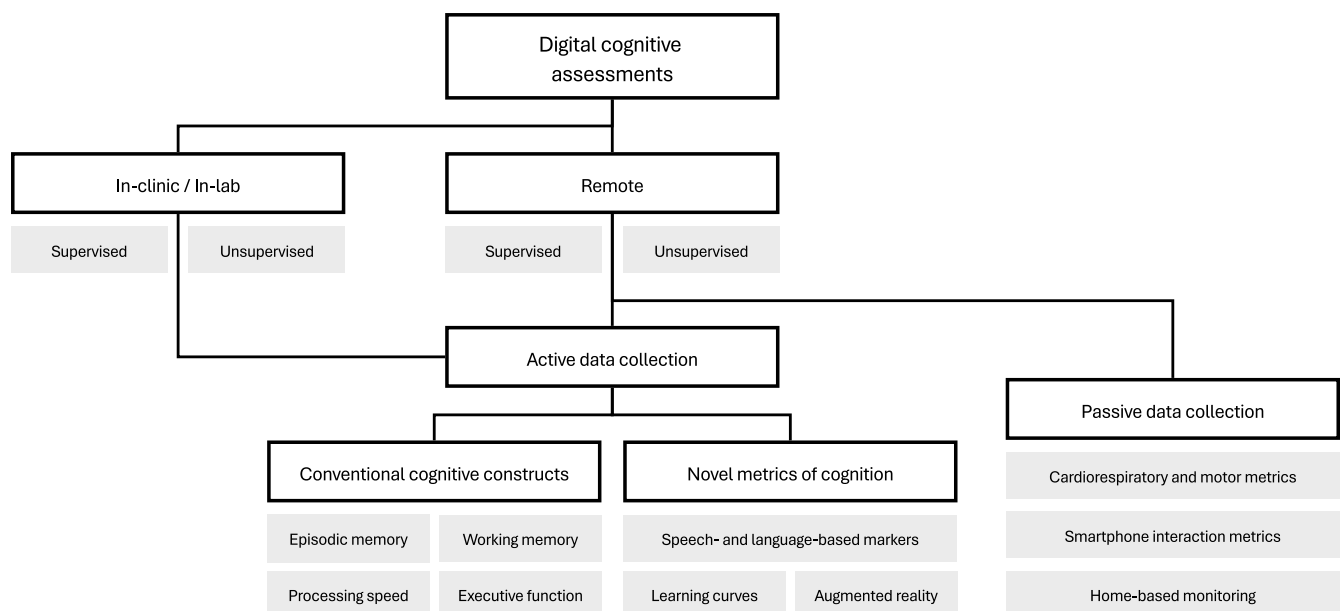


Fig. 1 | A non-exhaustive taxonomy of the types of digital cognitive assessments. A hierarchical diagram of categories under which digital cognitive assessments may fall in white, with subgroups and examples of these categories in gray. These

categories are not necessarily mutually exclusive. The current review focuses on those assessments that are remotely deployed without supervision, and which use active data collection to quantify cognitive function (i.e., the center of the diagram).

more reflective of individuals' everyday cognitive function, as it may be less associated with increased anxiety during testing⁵². On the other hand, unproctored testing introduces challenges related to environmental distractions, low effort, and malingering, as well as data fidelity (e.g., ref. 53). Building attention checks into tasks and asking explicitly if participants were distracted while completing the tasks may limit these effects. Researchers may also consider implementing algorithms to flag data patterns indicative of cheating, as well as integrating reliable participant authentication processes, data privacy policies permitting.

Another challenge relates to the level of digital literacy necessary for individuals to complete smart device-based testing without assistance, which can vary depending on the target population and tends to be lower in older samples⁵⁴. Additionally, access to a smart device and a dependable network connection may be a limitation, for example, when working with low-income populations or those in remote rural areas. According to a report published by the International Telecommunication Union⁵⁵, 67% of the global population uses the Internet. Notably, this statistic is highly variable across regions and income levels, and Internet access is particularly limited in Africa compared to other continents, as well as in rural areas and low-income countries⁵⁵. These statistics should be borne in mind when planning remote data collection and interpreting results, especially considering the strong role that both age and socioeconomic status play in neurodegenerative disease^{56,57}.

Finally, the underlying infrastructure and resources dedicated to data storage and handling will impose an upper limit on the scalability of any remote data collection, and the implementation of big data storage technologies⁵⁸ should be considered sooner rather than later. As such data storage and transfer systems are engineered, particular thought should be given to data privacy and protection, as individual cognitive health data is highly sensitive^{59,60}. Researchers should understand the risks as well as costs associated with data transfer and storage to make informed decisions that comply with relevant regulatory frameworks.

With these benefits and challenges in mind, as well as the current need for sensitive cognitive testing in preclinical AD, this scoping review will provide an overview of available remote and unsupervised digital cognitive assessments used in individuals with no clinically identified cognitive decline, but with biomarker evidence for elevated levels of A β , or A β and tau. We aim to present a necessary update of this rapidly evolving field since the publication of our previous review⁹, with a particular focus on remote and unsupervised assessments rather than digital cognitive assessments as a whole. Open questions in 2021 included whether such remote and unsupervised assessments are feasible in preclinical AD samples, as well as whether such measurements are reliable, especially longitudinally. We discuss progress in the field in this regard, evaluating feasibility based on rates of consent, enrollment, adherence, and compliance, as well as user experience reported by participants (also known as usability validity in the industry-oriented V3+ Framework⁶¹, which was developed as a common framework for digital health technologies measuring a wide range of metrics; <https://dataacc.dimesociety.org/v3/>), and report the reliability (between- and within-person, parallel-forms) of each tool. We evaluate construct validity (analytical validity in the V3+ Framework) based on associations with established neuropsychological assessments. Regarding the validity for various use cases in preclinical AD (known as clinical validity in the V3+ Framework), we describe whether they can accurately classify individuals with elevated A β (and tau) burden, whether they correlate with continuous measures of A β and tau, and whether they can predict future cognitive decline. To date, very few studies have used passively collected markers to characterize cognition in preclinical AD^{31,32}, therefore we focus on studies using actively collected digital assessments in the current review. Finally, since the feasibility, reliability, and validity of remote and unsupervised cognitive testing for use in preclinical AD have been established conceptually, we discuss the direction in which the field is quickly moving, addressing important future use cases, such as scalable case-finding, longitudinal monitoring, and individualized risk assessment, as well as what should be considered in the development of future cognitive tools.

Results

Literature screening

A total of 2688 records were found, 784 of which were excluded as duplicates or records in languages other than English. The remaining 1904 were screened, and 28 relevant reports were found (see Fig. 2 for details), two of which were previously included⁹. These and the 23 tools described in these records are listed in Table 1.

Remote capture of both conventional and novel metrics of cognition

Seventeen of the included tools capture the conventional cognitive constructs: Altoida¹⁴, Ambulatory Research in Cognition (ARC)^{13,37}, Boston Remote Assessment for Neurocognitive Health (BRANCH)⁶², Cambridge Neuropsychological Test Automated Battery (CANTAB)^{63,64}, Cogstate Brief Battery (CBB)^{7,63,65}, cCOG⁶⁴, Cognitron⁶⁶, the ki:elements Speech Biomarker for Cognition (ki: SB-C)^{67,68}, Mayo Test Drive (MTD)^{69,70}, MemTrax Memory Test⁶³, Mezurio³⁰, Mobile Monitoring of Cognitive Change (M2C2)⁷¹, neotiv^{44,46,64,66}, NeuroVocalix⁶⁴, NIH Mobile Toolbox⁶³, Oxford Cognitive Testing Portal (OCTAL)⁷², and the ReVeRe Word List Recall Test (RWLRT)⁶³. One of these tools (Altoida) uses an augmented reality-based cognitive task to collect multi-modal data (i.e., simultaneous capture of active and passive markers during a cognitive task), one (BRANCH) is used to quantify learning of repeated stimuli over days to months (i.e., learning curves)^{11,16}, four (ki: SB-C, Mezurio, NeuroVocalix, RWLRT) evaluate cognitive function using speech-based metrics, and two (ARC and M2C2) use ecological momentary assessment (EMA)-based paradigms⁷³ (EMA is a method of frequently sampling variables at random times of the day while subjects are in their natural environments to capture intra-individual variability; see refs. 73,74). Two other tools are used to quantify learning curves: Computerized Cognitive Composite (C3) Face-Name Associative Memory Exam (FNAME)^{10,75}, and Online Repeatable Cognitive Assessment—Language Learning Test (ORCA-LLT)¹². Finally, an additional four tools use speech-based metrics: Novoic^{76,77}, the Speech for Intelligent cognition change tracking and DETection of Alzheimer's Disease (SIDE-AD) online platform⁷⁸, Winterlight Assessment (WLA)¹⁵, and TapTalk⁷⁹, the last of which combines speech and motor function. Nine of these tools were discussed in the previous review either as in-person (e.g., CANTAB, NIH Toolbox, and Cogstate) or remote digital tools⁹, but had not yet been validated for remote use in preclinical AD (except BRANCH and ORCA-LLT^{10,12}).

High feasibility of remote and unsupervised data collection

These studies were generally highly feasible, as shown by rates of consent in studies, enrollment, adherence, compliance, and user experience reports. Three studies, all of which were recruited from ongoing studies, reported the consent rate of those individuals approached for participation: a year-long learning curves design with monthly assessments had a consent rate of 86%¹⁰, and a week-long measurement burst design^{39,40} with four daily measurements and had a consent rate of 87%³⁴. Finally, a longitudinal paradigm with biweekly assessments for 12 months also reported an 86% consent rate⁴⁶. However, the latter also reported that 24% of those who consented did not enroll in the app⁴⁶.

Adherence, generally measured as the average percentage of tests completed out of the complete study protocol, ranged from 74 to 93% for week-long protocols with up to four daily measurements^{11,13,15,37,71,76}, from 63 to 94% for an 8-week study¹⁹, and from 75 to 78% in year-long studies^{12,75}. Compliance, defined here as the completion of measurements as intended (e.g., acceptable data quality, successful attention checks), was generally excellent, with only 2 to 3% of data being unusable in both cross-sectional and longitudinal designs measuring conventional cognitive metrics as well as learning curves^{15,62,72,75}. However, 21% of data from the in-home augmented reality tasks were unusable due to technical issues, and another 32% of participants were unable to complete the in-home tasks either because their smartphone was not compatible or for other unspecified reasons¹⁴. Similarly, preliminary feasibility results from the current Alzheimer's

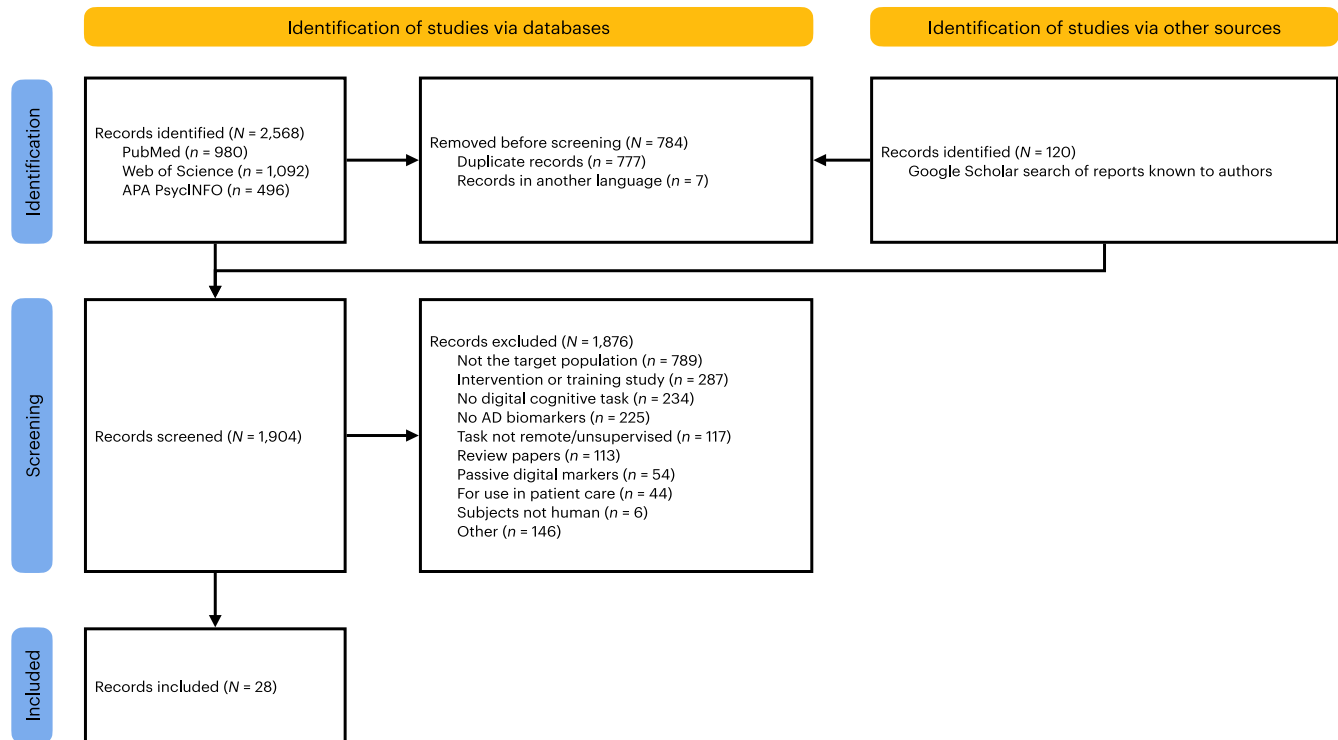


Fig. 2 | Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram detailing the screening of records. Number of records identified from which sources (databases and others), pre-screening of duplicates

and records in a language other than English, screening and exclusion of records (with reasons), and inclusion of records are detailed.

Disease Neuroimaging Initiative cohort (ADNI4) showed that only 54% of eligible individuals completed speech-based tasks⁸⁰.

Regarding user experience, in a cross-sectional paradigm, 16% of participants reported they had technical difficulties⁶², and in a week-long speech-based paradigm, 15% reported difficulties⁴⁷, while in a year-long design, 30% of participants reached out for technical support¹⁰. Tools were generally straightforward and enjoyable to use^{11,13,15,37,46,47,62,71,81}, with 65% of one sample reporting that they would “definitely” complete such sessions again³⁸. Additionally, a majority of participants indicated that they preferred remote testing to pen-and-paper testing (75%⁴⁶ and “most”³⁷).

Acceptable reliability of remote and unsupervised data collection

We also evaluated the reliability of each tool, insofar as this was reported (see Table 1 for specific statistics). Between-person reliability (i.e., variability across individuals, calculated with intraclass correlations [ICCs] according to ref. 82) was good for ARC and across multiple M2C2 assessments, with ICCs of 0.76 or greater^{13,37,83}. Parallel-forms reliability (i.e., reliability across alternate versions of the same task) was good for C3 FNAME (Cronbach’s $\alpha > 0.85$)²⁰, as well as for Novoic on average across versions (mean $\rho = 0.73$)⁷⁶; no other studies reported parallel-forms reliability. Test-retest reliability (i.e., precision of measurement within an individual) was moderate to excellent for most tasks (ICCs = 0.65 to 0.95, $r_s = 0.57$ to 0.81)^{21,37,46,62,84,85}, poor for Altoida (ICC = 0.48)¹⁴, and varied widely across WLA metrics (ICCs = -0.06 to 0.97)¹⁵, though most reached ICCs ≥ 0.50 across multiple assessments.

Construct and criterion validity of remote and unsupervised data collection

Finally, we examined the construct validity of each tool based on associations with in-person neuropsychological tests, as well as criterion validity for use in preclinical AD, based on associations with biomarkers of AD pathology (see Table 1).

A number of tools reported cross-sectional associations with established in-person cognitive measures. Those tools that reported associations with measures of global cognition (e.g., Preclinical Alzheimer’s Cognitive Composite [PACC]^{86,87} or similar composites) found correlations coefficients (r_s and ρ_s) between |0.53| and |0.70| and a standardized β estimate of 0.26^{14,37,46,62}. Two studies found associations between remote tasks and traditional neuropsychological tests measuring the same cognitive constructs ($r_s = |0.59|$ to $|0.61|$; $\beta_s = |0.22|$ to $|0.44|$)^{83,85}. Another study used speech-based metrics to predict PACC5 and found that predicted and measured PACC5 scores correlated ($r = 0.74$)⁷⁶. Additionally, some studies looked at the associations between remote task performance (cross-sectional and learning curves) and change in a cognitive composite (PACC), with change being retrospective, concurrent, or subsequent relative to the remote task administration ($r_s = 0.54$ to .69, $\beta_s = 0.27$ to 0.59)^{11,12,75}. Finally, one tool was able to distinguish those with an annual PACC score decline greater than 0.10 SD from non-decliners (area under the curve [AUC] = 0.91)⁷⁵.

Regarding associations with biomarkers of A β and/or tau pathology, a number of studies reported differences in performance between biomarker-negative and -positive groups^{10–16}. Most group differences had medium effect sizes (Cohen’s $d_s = 0.49$ to 0.66, Hedge’s $g_s = 0.43$ to 0.63, $\beta_s = |0.30|$ to $|0.40|$), with A β + (or A β +/-tau+) groups performing worse than A β – groups. One study found a small effect of time of day (Cohen’s $d = 0.19$), with individuals with elevated p-tau181/A β 42 levels performing worse in the evening¹³, while another found a very large effect of A β on learning curves across a year, with an A β + group showing dramatically slowed learning (Cohen’s $d = 2.22$)¹². A number of studies reported associations between performance on remote tasks and continuous measures of A β and/or tau^{10,12,37,46,62,70,72,75}, with correlation coefficients ranging from |0.11| to |0.34| and β_s ranging from |0.23| to |0.38|. Finally, some studies reported that remote task performance could distinguish between A β – and A β + individuals^{14,69,71,76} or A β /tau– and A β /tau+ individuals⁶⁹, with AUCs between 0.63 and 0.83.

Table 1 | Remote and unsupervised digital cognitive assessments for use in preclinical AD samples in alphabetical order

Tool	Report included	N included in analyses or ongoing study	Cognitive domain (task)	Reliability	Analytical and clinical validity according to associations with:	
					Neuropsychological tests	A β /tau
Altoida	Muurling et al., 2023 ¹⁴	56 (29 PET or CSF A β –, 27 A β +) + 65 ^b	Episodic memory (Augmented reality tasks) High-dimensional multi-modal data	Test-retest reliability: ICC = 0.48	Corr. with a cognitive composite (ρ = .56)	Lower scores in A β – than A β – (β \approx –0.40) ⁸ⁱ ; distin. A β – vs. A β – (AUC = 0.76)
	Wilks et al., 2021 ¹³	169 (81 CSF pTau:A β 42–, 32 pTau:A β 42+)	Episodic memory (Prices) Processing speed (Symbols) Working memory (Grids) EMA-based	Between-person reliability: ICCs > 0.80		Sundowning in CSF p-tau:A β 42+ group (Cohen's d = 0.19)
ARC	Nicosia et al., 2023 ³⁷	290 (212 with PET; 146 with CSF)		Between-person reliability: ICCs > 0.81; test-retest reliability: ICCs > 0.85	Corr. with a cognitive composite (r = –0.53) ^h	Corr. with A β (r = 0.26) and tau PET (0.11); CSF A β 42 (–0.23), t-tau (0.28), p-tau181 (0.25) ^h
	Papp et al., 2021 ⁶²	234 (144 with PiB; 129 with FTP)	Episodic memory (Categories, Face-Name-Occupation, Groceries, Signs) Processing Speed (Digit-Signs) Learning curves	Test-retest reliability: r = 0.81	Corr. with PACC (r = 0.62)	Corr. with cortical A β (r = –0.21) and entorhinal tau PET (–0.18)
BRANCH	Papp et al., 2024 ¹¹	164 (128 PET A β –, 36 A β +) + 65 ^b		Test-retest reliability: ICC = 0.94 ²¹	Corr. with subsequent decline on PACC (r = 0.54)	Reduced learning curves in A β – (Cohen's d = 0.49)
	Jutten et al., 2025 ¹⁶	167 (106 PET A β –/tau–, 46 A β –/tau–, 15 A β –/tau+) + 65 ^b				Reduced learning curves in A β –/tau+ (β s = –0.47 to –0.61), and all A β – (–0.24 to –0.58)
CANTAB	Weiner, Aaronson, et al., 2023 ⁶³	Brain Health Registry	Episodic memory (Delayed Matching to Sample, Paired Associates Learning ^d , Pattern Recognition Memory, Verbal Paired Associates, Verbal Recognition Memory) Working memory (Spatial Working Memory, Digit Span, N-Back, Spatial Span)		Awaiting validation in preclinical AD	
	Malzbender et al., 2024 ⁶⁴	AD-RIDDLE	Executive function (Cambridge Gambling Task, Digit Symbol Substitution Task, Intra-Extra Dimensional Set Shift, Multitasking Test, One-Touch Stockings of Cambridge, Stockings of Cambridge, Stop Signal Task, Match to Sample Visual Search, Rapid Visual Information Processing) Psychomotor (Motor Screening Task, Reaction Time, Adaptive Tracking Task) Emotion and social (Emotional Bias Task, Emotion Recognition Task)		Awaiting validation in preclinical AD	
CBB	Langford et al., 2020 ⁷	4486 (3163 PET A β –, 1323 A β +) + 65 ^b	Episodic memory (One Card Learning) Working memory (One-Back) Executive function (Identification) Psychomotor (Detection)	Test-retest reliability: ICCs = 0.90 to 0.95 ⁹⁸		Distin. A β – vs. A β – (AUCs = 0.60 to 0.73)
	Kaye et al., 2021 ⁶⁵	TRC-PAD			Awaiting further validation in preclinical AD	
	Weiner, Aaronson, et al., 2023 ⁶³	Brain Health Registry			Awaiting further validation in preclinical AD	
cCOG	Malzbender et al., 2024 ⁶⁴	AD-RIDDLE	Episodic memory (Learning task, Recall task, Recognition task) Processing speed (Trail Making Tests) Psychomotor (Reaction tests)		Awaiting validation in preclinical AD	
Cognitron	Leuzy et al., 2024 ⁶⁶	REAL AD	Episodic memory (Card Pairs, Mallas Memory Short, Paired Associates Learning) Working memory (Digit Span, Reverse Digit Span, Number Location Pairs, Picture Completion, Spatial Span) Processing speed (Blocks, 2D Manipulations, Choice Reaction Time, Trail Making Test) Executive function (Selective Attention, Target Detection, Switching Stroop, Stop Change Task, Tower of London) Fluid intelligence (Verbal Reasoning, Information Sampling, Odd One Out) Psychomotor (Motor Control, Simple Reaction Time) Mathematics (Balloons)		Awaiting validation in preclinical AD	

Table 1 (continued) | Remote and unsupervised digital cognitive assessments for use in preclinical AD samples in alphabetical order

Tool	Report included	N included in analyses or ongoing study	Cognitive domain (task)	Reliability	Analytical and clinical validity according to associations with:	
					Neuropsychological tests	Aβ/tau
C3 FNAME	Samaroo et al., 2020 ^{10,a}	94 (69 PET Aβ−, 25 Aβ+; 84 with FTP)	Episodic memory (Face Name Associative Memory Exam) Learning curves	Parallel-forms reliability: Cronbach's $\alpha > 0.85^{20}$		Reduced learning curves in Aβ+ (Cohen's $d = 0.66$), corr. with tau ($r = -0.22$)
	Jutten et al., 2022 ⁷⁵	114 (81 PET Aβ−, 33 Aβ+; 105 with FTP)			Corr. with concurrent PACC5 decline ($r = 0.69$); distin. PACC5 decliners (AUC = 0.91)	Corr. with global Aβ ($r = -0.20$), entorhinal tau (-0.38), and inf.-temp. tau PET (-0.23)
kic SB-C	Gregory et al., 2022 ⁶⁷	SPeAk	Structured speech-based tasks	Test-retest reliability: $rs = 0.57$ to 0.72^{84}	Awaiting validation in preclinical AD	
	König et al., 2023 ⁶⁸	PROSPECT-AD			Awaiting validation in preclinical AD	
MemTrax Memory Test	Weiner, Aaronson, et al., 2023 ⁶³	Brain Health Registry	Episodic memory		Awaiting validation in preclinical AD	
Mezurio	Muurling et al., 2021 ³⁰	RADAR-AD	Episodic memory (Gallery Game) Executive function (Tilt Task) Structured speech-based task (Story Time)		Awaiting validation in preclinical AD	
MTD	Stricker et al., 2024 ⁶⁹	353 (228 PET Aβ−, 125 Aβ+; 250 with FTP)	Episodic memory (Stricker Learning Span) Processing speed (Symbols)	Test-retest reliability: ICCs $> 0.71^{65}$	Corr. with in-person NPTs of same construct ($rs = 0.26 $ to $ 0.51 ^{65}$)	Distin. Aβ− vs. Aβ+ (AUCs = 0.63 to 0.77), Aβ−/tau − vs. Aβ+/tau+ (0.67 to 0.83)
	Boots et al., 2024 ⁷⁰	684 (670 with PiB; 667 with FTP)			Corr. with PACC ($r = 0.68$); corr. with in-person NPTs of same construct ($rs = 0.59 $ to $ 0.61 $)	Corr. with Aβ ($\rho = -0.24$), entorhinal tau (-0.23), and global tau PET (-0.21)
M2C2	Thompson et al., 2023 ⁷¹	69 (44 PET Aβ−, 25 Aβ+)	Episodic memory (Prices) Processing speed (Symbol Match) Working memory (Color Shapes) EMA-based	Between-person reliability: ICCs = 0.25 to 0.97 ^{63,e}	Corr. with in-person NPTs of same construct ($\beta s = 0.22 $ to $ 0.44 ^{63}$)	Distin. Aβ− vs. Aβ+ (AUCs = 0.73 to 0.77)
neotiv	Berron, Olsson, et al., 2024 ⁴⁶	100 (58 PET Aβ−, 42 Aβ+; 100 with FTM)	Episodic memory (Mnemonic Discrimination Task, Object-in-Room Recall, Complex Scene Recognition)	Test-retest reliability: ICCs = 0.65 to 0.83	Corr. with PACC ($rs = 0.62$ to 0.70); pred. concurrent PACC decline ($\beta s \approx 0.41$ to 0.59)	Pred. Aβ in precuneus ($\beta s = 0.23$ to 0.38) and tau in the MTL (0.20 to 0.40) ⁱ
	Berron, Olsson, et al., 2024 ⁴⁶	BioFINDER-2			Awaiting further validation in preclinical AD	
	Berron, Glanz, et al., 2024 ⁴⁴	DELCODE, WRAP			Awaiting further validation in preclinical AD	
	Malzbender et al., 2024 ⁶⁴	AD-RIDDLE			Awaiting further validation in preclinical AD	
	Leuzy et al., 2024 ⁴⁶	REAL AD			Awaiting further validation in preclinical AD	
NeuroVocalix	Malzbender et al., 2024 ⁶⁴	AD-RIDDLE	Speech-based episodic memory		Awaiting validation in preclinical AD	
NIH Mobile Toolbox	Weiner, Aaronson, et al., 2023 ⁶³	Brain Health Registry	Episodic memory (Faces and Names with Delay, Arranging Pictures) Working memory (Sequences) Processing speed (Number-Symbol Match) Executive function (Arrow Matching, Shape-Color Sorting) Language (Spelling, Word Meaning)		Awaiting validation in preclinical AD	

Table 1 (continued) | Remote and unsupervised digital cognitive assessments for use in preclinical AD samples in alphabetical order

Tool	Report included	N included in analyses or ongoing study	Cognitive domain (task)	Reliability	Analytical and clinical validity according to associations with:	
					Neuropsychological tests	A β /tau
Novoic	Fristed et al., 2022 ⁷⁶	115 (59 PET or CSF A β −, 56 A β +)	Speech-based storytelling	Parallel-forms reliability: ρ s = 0.39 to .85 ^{47,f}	Predicted PACC corr. with measured PACC ($r = 0.74$)	Distin. A β − vs. A β + (AUC = 0.74)
	Weiner, Veitch, et al., 2023 ⁷⁷	ADNI4			Awaiting further validation in preclinical AD	
OCTAL	Toniolo et al., 2024 ⁷²	99 (99 with plasma) + 352 ^c	Episodic memory (Object-in-Scene Memory Task, Rey-Osterrieth Complex Figure) Working memory (Oxford Memory Test, Freestyle Corsi Block Task) Processing speed (Digit Symbol Substitution Test, Trail Making Test)			Corr. with plasma p-tau181 ($r = 0.34$); pred. plasma p-tau181 ^l
ORCA-LLT	Lim et al., 2020 ^{12,a}	80 (42 PET A β −, 38 A β +)	Learning curves		Pred. baseline EM ($\beta = 0.26$); pred. retrospective EM change ($\beta = 0.27$)	Reduced learning curves in A β + (Cohen's $d = 2.22$); pred. A β PET ($\beta = -0.23$)
RWLRT	Weiner, Aaronson, et al., 2023 ⁶³	Brain Health Registry	Speech-based episodic memory		Awaiting validation in preclinical AD	
SIDE-AD Platform	Saunders et al., 2024 ⁷⁸	SIDE-AD	Unstructured speech-based tasks		Awaiting validation in preclinical AD	
TapTalk	Alty et al., 2024 ⁷⁹	Validation studies of TapTalk	Speech- and motor-based task		Awaiting validation in preclinical AD	
WLA	van den Berg et al., 2024 ¹⁵	50 (27 PET A β −, 23 A β +)	Structured and unstructured speech tasks	Test-retest reliability: ICCs = −0.06 to 0.97 ^g		More pauses in A β + than A β − (β s ≈ 0.30 to 0.34) ^h

ARC Ambulatory Research in Cognition, *BRANCH* Boston Remote Assessment for Neurocognitive Health, *C3 FNAME* Computerized Cognitive Composite Face-Name Associative Memory Exam, *CANTAB* Cambridge Neuropsychological Test Automated Battery, *CBB* Cogstate Brief Battery, *ki-e SB-C* ki-elements Speech Biomarker for Cognition, *M2C2* Mobile Monitoring of Cognitive Change, *MTD* Mayo Test Drive, *NIH* National Institutes of Health, *OCTAL* Oxford Cognitive Testing Portal, *ORCA-LLT* Online Repeatable Cognitive Assessment—Language Learning Test, *RWLRT* ReVeRe Word List Recall Test, *SIDE-AD* Speech for Intelligent cognition change tracking and DEtection of Alzheimer's disease, *WLA* Winterlight Assessment; *PET* positron emission tomography, *CSF* cerebrospinal fluid, *PiB* Pittsburgh compound B, *FTP* flortaucipir, *FTM* flutemetamol, *EMA* ecological momentary assessment, *PACC* Preclinical Alzheimer's Cognitive Composite, *NPTs* neuropsychological tests, *corr.* correlated, *pred.* predicted, *distin.* distinguished.

^aIncluded in ref. 9

^b $n = 65$ did not complete remote testing.

^c $n = 352$ were included as an online sample with no biomarkers.

^dOnly Paired Associates Learning was announced to be used in association with the Brain Health Registry.

^eBetween-person reliability was greatly improved for the M2C2 when aggregating multiple measurements (≥ 0.76).

^fNovoic tested 153 correlations across test versions, and average ρ s reached 0.73.

^gWLA tested 186 ICCs; reliability reached ICCs of ≥ 0.50 across two or more assessments.

^hHigher ARC score indicates worse performance, thus the correlations reported here are in the expected direction.

ⁱUnstandardized estimates originally reported were standardized using the standard errors or 95% confidence intervals and N s reported.

^jOnly p values were reported for the models, including individual metrics. The best model to predict p-tau181 combined remote and in-person metrics, as well as A β 42/40, adjusted $R^2 = 0.50$.

For samples including groups other than preclinical AD, effect sizes are reported for the findings pertaining to the preclinical AD sub-sample.

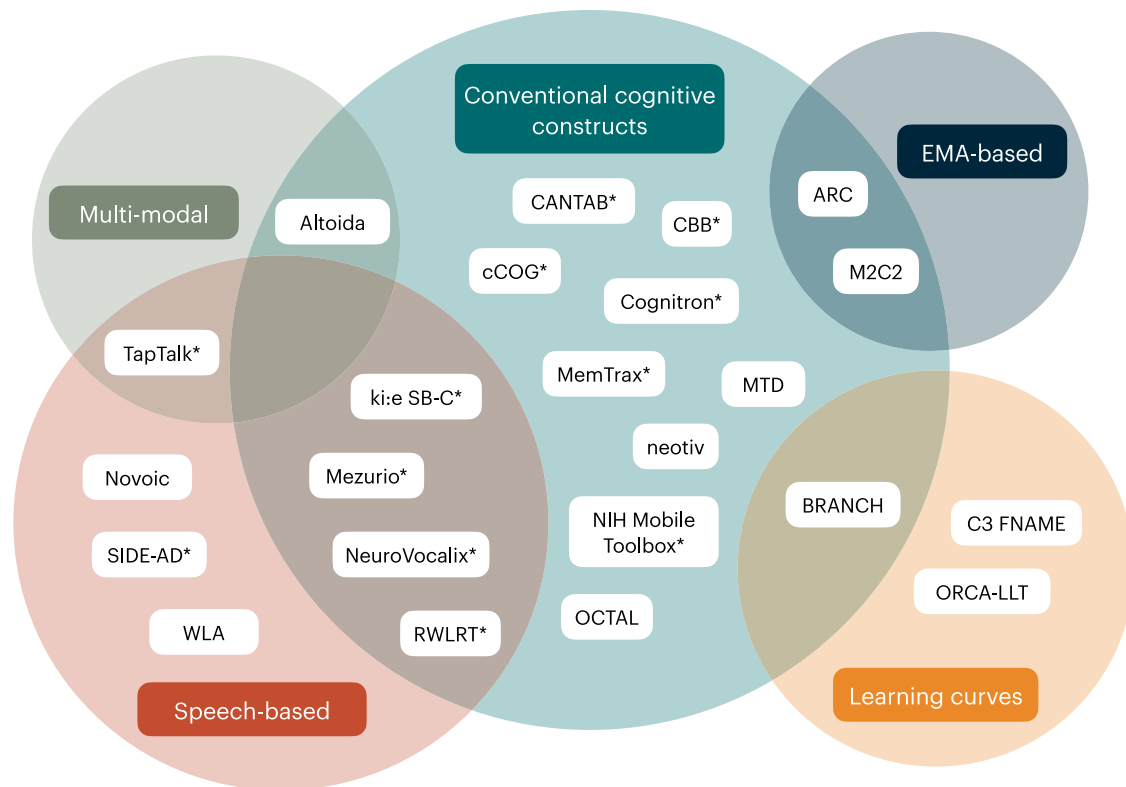


Fig. 3 | Venn diagram of the tools included in the current review based on the type of cognitive metrics they quantify. Tools identified in the current scoping review were categorized according to their methodology and metrics that they quantified. Tools measuring conventional cognitive constructs are shown in the teal circle; tools using multi-modal data collection are shown in the green circle; tools capturing speech-based metrics are shown in the red circle; tools using EMA-based protocols are shown in the dark blue circle; and tools quantifying learning curves are shown in the yellow circle. Some tools are considered to belong to multiple categories. EMA ecological momentary assessment, ARC Ambulatory Research in cognition, BRANCH Boston Remote Assessment for Neurocognitive Health, C3 FNAME

Computerized Cognitive Composite Face-Name Associative Memory Exam, CANTAB Cambridge Neuropsychological Test Automated Battery, CBB Cogstate Brief Battery, ki:e SB-C ki:elements Speech Biomarker for Cognition, M2C2 Mobile Monitoring of Cognitive Change, MTD Mayo Test Drive, NIH National Institutes of Health, OCTAL Oxford Cognitive Testing Portal, ORCA-LLT Online Repeatable Cognitive Assessment—Language Learning Test, RWLRT ReVeRe Word List Recall Test, SIDE-AD Speech for Intelligent cognition change tracking and DEtection of Alzheimer's disease, WLA Winterlight Assessment. *Tools awaiting validation for use in preclinical AD samples.

Remote and unsupervised tools awaiting validation in preclinical AD

A number of tools included in the current review have yet to be validated for use in preclinical AD, or are undergoing further validation for additional use cases, but their deployment in relevant samples has been announced (see Table 1). These tools are included in moderately sized to large studies collecting longitudinal data (e.g., ADNI4, BioFINDER-2, DELCODE, PROSPECT-AD, RADAR-AD, REAL AD, SIDE-AD, SPeAk, WRAP), in registries that collaborate with such studies (e.g., Brain Health Registry), and in large collaborative efforts to inform the research and treatment of AD (e.g., AD-RIDDLE). They include tools that capture conventional cognitive constructs as well as novel metrics of cognitive function, predominantly speech-based markers. All of these tools are deployed in samples that include biomarker-characterized cognitively healthy older adults, and results regarding the validation for their use in preclinical AD are expected.

Discussion

In this review, we focused on remote and unsupervised digital cognitive assessment tools that capture both conventional cognitive constructs and novel metrics with active data collection to characterize and detect subtle cognitive decline in preclinical AD. We defined preclinical AD as the absence of clinically established cognitive impairment in the presence of markers of A β , and sometimes tau, pathology. We found that remote and unsupervised cognitive assessments generally have good feasibility and validity for use in preclinical AD, and that the field is quickly moving

forward with larger samples and longitudinal studies to address relevant use cases in both clinical trials and healthcare contexts (see Fig. 3).

At the time our previous review was published⁹, the question of whether it is feasible to remotely deploy digital cognitive tools without supervision in preclinical AD samples was still open. Among the studies covered in the current review (two that were previously included^{10,12}) we found that rates of consenting into studies (albeit from other ongoing studies), adherence (i.e., how many measurements participants completed), and compliance (i.e., how many measurements participants completed as intended) were impressive. Across study designs ranging from one week to one year, consent rates ranged from 86 to 87%, adherence rates from 63 to 93%, and compliance rates from 97 to 98%. This is in stark contrast to a previous report of the median participant retention in digital health studies being only 5.5 days out of 12 weeks⁸⁸. Another indicator of usability was the generally positive user feedback collected with user experience surveys. However, it is important to note that many digital assessment studies included here recruited from existing longitudinal study cohorts; participants in these studies typically agree to complete fluid biomarker acquisition, neuroimaging, and regular in-clinic assessments and therefore do not represent the general population. Adherence reported in the studies included in this review may, therefore, be inflated compared to the expected adherence to such paradigms among real-world samples.

One potential bottleneck to participant retention may be registration in digital apps after consenting to a study. For example, out of those who consented to participate in a year-long study, only 64% actually registered in

the smartphone app and completed at least one task⁴⁶. Similarly, of those individuals who met ADNI4 inclusion criteria, only 54% completed speech-based remote tasks⁸⁰. One factor that may ameliorate participant drop-out is the technical support available; while 15 to 16% of participants in cross-sectional or week-long studies reported technical difficulties^{47,62}, 30% of participants in a year-long longitudinal study reached out to the study team for technical support¹⁰. Additionally, as the tasks become more technologically demanding, feasibility may become limited, as was the case for the augmented reality-based tasks, during which 21% of participants had technical issues precluding the use of their data, and another 32% could not complete the tasks because of smartphone incompatibility or other unnamed reasons¹⁴. Indeed, when issues with remote technology were quantified (from 0 to 1, 0 = no problems) for both active cognitive tools included in RADAR-AD, the same participants had fewer technology-based problems with the app used for administering conventional cognitive tasks (0.31) than with the augmented reality-based app (0.60)¹⁹.

As more and more large longitudinal studies implement remote data collection, special attention should be paid to participant enrollment and retention, especially when recruitment is done remotely and outside of Western, Educated, Industrial, Rich, and Democratic (WEIRD) populations. One way of boosting retention, for example, is by having participants engage with the digital tool immediately after consenting to a study⁸⁹. Having technical support available to participants, especially those with lower digital literacy⁵⁴, may also alleviate both issues of limited enrollment as well as prohibitive technical difficulties during tasks.

Another open point in the previous review pertained to the reliability of remotely assessed cognitive metrics⁹. The reporting of reliability varied across studies, but reported reliability was generally good. Three studies reported between-person reliability (i.e., the ability of the test to differentiate individuals), two using ARC and one using M2C2^{13,37,83}; between-person reliability of ARC was high, as well as across multiple sessions of M2C2. Parallel-forms reliability was only reported by two studies: C3 FNAME showed good reliability across alternate versions²⁰, while Novoic tasks had moderate to excellent reliability across versions⁴⁷. Test-retest reliability (i.e., the reliability of a test within an individual across multiple assessments) was more widely reported, with moderate to excellent reliability of five tools capturing conventional cognitive constructs as well as learning curves^{21,37,46,62,85}. The test-retest reliability for the speech-based WLA varied depending on the metric and the number of assessments across which performance was averaged¹⁵. For both neotiv and WLA, test-retest reliability improved when aggregating across multiple measurements^{15,46}. Altoida was the only tool that reported poor retest reliability¹⁴, which was acceptable for Android users, ICC = 0.70, but very poor for iOS users, ICC = 0.33. Authors speculated that there may have been version effects, the sample may have been too small to accurately capture test-retest reliability ($n = 43$), or participants may have received help during some measurements but not others¹⁴.

Overall, reliability was generally favorable, suggesting that cognitive testing can reliably separate individuals and also precisely capture cognition within the same individual across multiple measurements, even when done remotely and without supervision. As longitudinal data collection and the quantification of cognitive changes become the goal of more and more studies, the reliability of assessment tools, especially test-retest reliability, should be carefully assessed in all target populations, and reasons for poor reliability should be corrected to ensure that the capture of subtle change is not obscured by random noise.

Regarding construct validity, the cross-sectional associations between remotely administered digital cognitive tools and in-person neuropsychological batteries indicate that these remote tools generally map onto cognitive constructs that are important in the clinical quantification of cognitive decline. Additionally, three studies found associations between remote task performance (including learning curves) and change in PACC score, indicating remote task performance may be useful as a prognostic tool to predict cognitive decline. In general, associations were moderate to strong, however no correlations greater than 0.70 were found. As a benchmark, we looked at

digital tests that had been administered remotely as well as in-person: Altoida performance was correlated across at-home and in-clinic administration with a Spearman's ρ of 0.57¹⁴, as was C3 (composite score across all tasks including FNAME) with an r of 0.70²⁰, while a remote version of the neotiv Mnemonic Discrimination Task correlated with a similar in-scanner version with an r of 0.66⁴⁶. This suggests that remote and in-clinic assessments, even when using nearly identical measures, may only be correlated to a limited degree.

Regarding these moderate correlations, Stricker and colleagues⁶⁹ argue that correlating novel measures of cognition with established measures is also not a foolproof method of validation, since the existing measures of cognition, although extensively validated, are themselves imperfect⁹⁰. Many existing neuropsychological tests were developed to quantify cognitive impairment in symptomatic individuals years or even decades before the emergence of reliable in vivo biomarkers of AD and neuroimaging^{90–92}. The constructs that traditional neuropsychological assessments capture may, therefore, be (1) insensitive to changes in non-symptomatic individuals (e.g., cognitively healthy individuals may reach a ceiling) and (2) outdated in terms of our current understanding of the biological progression of AD. In comparison, the development of novel tests of cognition in AD can be tailored to capture subtle changes as well as be neuroanatomically informed (e.g., neotiv⁴⁶). Additionally, the target population should again be taken into account here; for example, there may be reduced digital familiarity among older adults⁵⁴, further limiting correlations between in-person and remote tests. Focusing on other ways of establishing clinical validity, such as using known groups (e.g., cognitively unimpaired A β – and A β + groups⁶⁹), may lead to greater precision in measuring behaviors that are meaningful for preclinical AD.

Necessarily, all tools included in this review were used to find associations between cognition and markers of A β and/or tau pathology. Most studies used positron emission tomography (PET)-characterized A β and tau burden^{10–12,14–16,37,46,62,69,71,75,76}, while some used cerebrospinal fluid (CSF) markers, either as the main markers of interest or as a substitute in the case of missing PET scans^{37,76}. Most studies reported known-group validity, comparing A β – to A β + groups and finding poorer performance on remote cognitive tests among cognitively healthy individuals with elevated A β burden, showing that many of these assessments are sensitive to the subtle cognitive changes in preclinical AD. Associations between continuous measures of A β and/or tau and remotely assessed cognitive were also found, suggesting that cognitive performance as captured by these remote and unsupervised tests progressively worsens as AD pathology accumulates in the brain. Altogether, these findings support the claim that remote and unsupervised cognitive assessments can be used to quantify cognitive functions that are affected by AD pathology, even before clinical symptoms emerge.

So far, one study has looked at the associations between blood plasma markers, specifically p-tau181 and A β 42/40, and cognitive performance on remote digital tasks⁷², which included both cognitively unimpaired individuals as well as those with clinically characterized AD. They found that plasma p-tau181 correlated moderately with a number of digital cognitive markers, while plasma A β 42/40 only showed weak correlations (correlation coefficients for the whole sample were not reported in ref. 72). Within cognitively unimpaired individuals, a correlation was found between plasma p-tau181 and the *localization time* metric on a “What is where?” task (i.e., associative memory; $r = 0.34$), and no correlations with plasma A β 42/40 were observed⁷². Another study combined performance on a remote task with plasma p-tau217 to predict a future decline in PACC score in a mostly cognitively unimpaired sample ($\beta_s \approx 0.41$ to 0.59)⁴⁶. Notably, compared to a model including only demographic factors and plasma p-tau217, a model also including digital cognitive markers explained 7% more variance, with a significant improvement in model fit (difference in Akaike information criteria = -26.430)⁴⁶. These initial findings suggest that blood plasma markers, particularly p-tau markers, and digital cognitive markers may be used in conjunction as scalable and accessible tools to screen for AD in its earliest stages, reducing the need for expensive and invasive procedures such

as PET scans, lumbar punctures, and extensive in-person cognitive batteries for individuals with minimal cognitive complaints.

Moving forward, many remote and unsupervised tools are currently undergoing validation for use in longitudinal contexts related to preclinical AD, some with the explicit goal of recruiting more diverse samples (e.g., ADNI4). Once this is achieved, the monitoring of cognitive decline for other use cases, such as safety monitoring and neurocognitive endpoints in clinical trials⁸, will be greatly facilitated. These efforts are especially timely given the recent update to the US Food and Drug Administration's guidelines regarding drug trials in preclinical AD, in which cognitive endpoints may be sufficient in trials including participants with no detectable cognitive or functional impairment⁶.

Other use cases where remote and unsupervised cognitive assessment may play an integral role in the future include the detection of meaningful cognitive changes within individuals, with the goals of at-scale case finding for both clinical trials (e.g., inclusion into studies) as well as healthcare contexts (e.g., diagnostic support), and individualized prognoses (i.e., who is at risk for cognitive decline in the future), in addition to the longitudinal monitoring of cognition already discussed.

The use of remote tools for case finding has already been established in small samples. Four studies included here showed that cognitive function captured by remote tools could distinguish between cognitively healthy individuals with and without elevated A β burden: two tasks that used conventional cognitive tests, the MTD Stricker Learning Span (verbal learning, which also distinguished between A β /tau− and A β /tau+)⁶⁹ and the M2C2 tasks (the associative memory task performed best)⁷¹, Altdora, which used an augmented reality-based task¹⁴, and Novoic, which employed speech-based tasks⁷⁶. The best-performing task achieved an AUC of 0.77 (0.83 for A β /tau− vs. A β /tau+), and while this may not be considered sensitive enough for use as a stand-alone diagnostic instrument, it may be sufficient for use in screening for further confirmatory diagnostic testing, for example, with CSF or PET, or even blood-based biomarkers. Indeed, screening and recruitment for ADNI4 is already facilitated by the use of remote digital cognitive assessments⁷⁷, showing their value as a screening tool for inclusion in AD research.

The abovementioned longitudinal studies will also be a valuable resource in validating the use of remote and unsupervised cognitive assessments for individualized prognostics. Three studies included in this review reported associations between cross-sectional performance on remote tasks or learning curves and concurrent and/or subsequent change in traditional neuropsychological test scores on the group level: performance on the neotiv Mnemonic Discrimination Task and Object-in-Room Recall, especially in combination with plasma p-tau217 and demographic factors, predicted PACC5 decline over up to 5 years⁴⁶, while learning curves captured with C3 FNAME and BRANCH showed associations with PACC change over one year^{11,75}. This suggests that remote cognitive assessments, perhaps in combination with less invasive blood-based biomarkers, may be informative for prognoses in terms of expected future cognitive decline in AD. Future research may seek to validate remote cognitive assessments in terms of how well they can discriminate between individuals who will decline in the future and those whose cognition will remain relatively stable (i.e., individual risk assessment).

The clinical validation of remote and unsupervised cognitive assessment tools for individualized prognostics would be a meaningful milestone for patients and caregivers in particular. The expected onset of clinical and functional decline or the average time until noticeable cognitive symptoms typically emerge or until independent living may become difficult can be valuable information for those affected by neurodegenerative diseases⁹³.

Regarding the practical use of current tools, as well as the development of future tools, we found a total of 23 tools that are used for remote and unsupervised data collection in preclinical AD, some of which measured overlapping constructs. Researchers looking to develop new tools to detect early changes in cognition due to AD pathology should review existing tools and their validated use cases to determine whether the development of a new tool is necessary, given the associated monetary costs as well as researcher and

participant/patient burden. As a reference for existing digital health technologies used for cognitive assessment as well as other clinically meaningful outcomes and predictors, such as sleep and physical activity, the Digital Health Measurement Collaborative Community (DATAcc) by the Digital Medicine Society (DiMe) has compiled a list of validated digital health tools in the Library of Digital Measurement Products⁹⁴ (<https://datacc.dimesociety.org/digital-measurement-library/>). As of January 20, 2025, all the tools in Table 1, except OCTAL and ORCA-LLT can be found in the library, and new digital tools are added regularly. The question of which tool is the “best” will depend heavily on the research question and/or use case. While a head-to-head comparison of existing tools is outside of the scope of the current review, it is an important goal to empirically establish which specific tools researchers and clinicians should use, depending on the respective use case, and indeed studies in this direction are already underway (e.g., ref. 64).

Any novel tools, especially those developing new markers of cognitive function, should also be subject to rigorous testing of feasibility, reliability, and validity (e.g., following established psychometric procedures and/or the V3+ Framework⁶¹). In general, tools for the remote and unsupervised assessment of cognitive function should have high compliance and usability, especially within the target population. They should reliably measure cognitive function across multiple administrations within the same person and, for use cases requiring longitudinal data collection, be relatively resilient to retest effects (unless they specifically aim to quantify practice effects). And finally, they should show associations, cross-sectional and/or longitudinal depending on the use case, with established measures of the same construct, as well as relevant validity for the use case of interest (e.g., case-finding, prognosis, monitoring).

The current scoping review has some limitations that should be acknowledged. Due to the large volume of records found upon the first search, the search process was not continued iteratively⁹⁵. This may have resulted in missed reports, though given the large amount of records screened and the nascency of the field, this is unlikely. Additionally, only reports in English were included, excluding reports in French, Chinese, German, Portuguese, and Spanish. Future reviews may consider screening reports in all languages for better inclusion of international findings.

In summary, this scoping review identified 28 papers reporting on 23 digital tools for the remote and unsupervised assessment of cognition in preclinical AD. We provided updates to open questions posed by Öhman and colleagues⁹, determining that remote studies of cognition in healthy older adults are largely feasible, with certain restrictions to usability, and that the data collected with such tools are generally reliable, opening the door for the use of such tools longitudinally. Finally, validity has been conceptually established for these tools with respect to their use in preclinical AD and should continue to be evaluated as these tools are implemented in new contexts of use. Currently, studies deploying remote cognitive assessment tools are focused on acquiring larger, more diverse samples over longer periods of time to validate the use of such tools for longitudinal monitoring of cognition. Future goals include exploring how remote and unsupervised digital tools can be used for case-finding on a scalable level—efforts in this regard are already being made (e.g., ADNI4), individualized prognostics and risk assessment—especially as it pertains to those affected by AD, and longitudinal monitoring of subtle changes in the earliest stages of AD. (Fig. 3).

Methods

Search strategy and selection criteria

An initial literature search was performed on September 12, 2023, using PubMed, Web of Science, and APA PsycINFO using terms including digital, remote, unsupervised, smartphone, cognition, Alzheimer's, and dementia (see Supplementary Methods for exact search terms) with no limitation according to date. Publications already known to the authors were also included. Literature searches were repeated on March 8, 2024, September 9, 2024, and January 14, 2025. All records were uploaded onto Rayyan, a web-based literature screening tool⁹⁶. Duplicates were algorithmically identified, then manually checked and excluded, as were records in languages other than English. The remaining titles and abstracts were screened by SEP and

FÖ; ambiguously relevant reports were flagged and reviewed by both authors. For each paper that was excluded, primary reasons for exclusion were recorded.

Peer-reviewed research articles and planned study reports as well as preprints were included if they used self-administered and remote assessments (e.g., no supervision via phone or video call) and if they reported findings related to preclinical AD (i.e., included measures of A β and/or tau pathology in cognitively healthy older adults). We included active cognitive tests (i.e., actively completed by participants), excluding those that only measured passive biomarkers (e.g., gait, keystrokes). Studies including other diseases (e.g., cancers, cardiovascular or other neurodegenerative diseases, psychiatric conditions) were excluded, as were studies using the cognitive assessment tool as an intervention.

Data extraction

Author(s), year of publication, sample size and group allocation (if applicable), name of cohort or ongoing study, names and types of cognitive tasks available, adherence and compliance metrics, usability metrics, and reliability metrics were extracted from each of the selected papers. If certain feasibility or psychometric information about the tools was not reported in the included articles, this was gathered from other papers using the same tools, sometimes in separate samples, found using Google Scholar. Associations with in-person neuropsychiatric assessments and measures of AD biomarkers were also charted. Additionally, if unstandardized estimates were reported in the original papers, these estimates were standardized using standard errors or 95% confidence intervals and Ns to be able to compare effect sizes; these calculations were not done with the original data and are considered approximations.

This manuscript was prepared according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist⁹⁷, which can be found in the Supplementary Information.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study. Literature search results can be made available upon request.

Received: 25 September 2024; Accepted: 25 March 2025;

Published online: 10 May 2025

References

- Jack, C. R. et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's association workgroup. *Alzheimers Dement.* **20**, 5143–5169 (2024).
- Jack, C. R. et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* **14**, 535–562 (2018).
- Dubbelman, M. A. et al. Cognitive and functional change over time in cognitively healthy individuals according to Alzheimer disease biomarker-defined subgroups. *Neurology* **102**, e207978 (2024).
- Jutten, R. J. et al. Why a clinical trial is as good as its outcome measure: a framework for the selection and use of cognitive outcome measures for clinical trials of Alzheimer's disease. *Alzheimers Dement.* **19**, 708–720 (2023).
- Jessen, F. et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement.* **10**, 844–852 (2014).
- Food and Drug Administration. Early Alzheimer's disease: developing drugs for treatment. Guidance for industry. (2024).
- Langford, O. et al. Predicting amyloid burden to accelerate recruitment of secondary prevention clinical trials. *J. Prev. Alzheimers Dis.* **7**, 213–218 (2020).
- Hampel, H. et al. Designing the next-generation clinical care pathway for Alzheimer's disease. *Nat. Aging* **2**, 692–703 (2022).
- Öhman, F., Hassenstab, J., Berron, D., Schöll, M. & Papp, K. V. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **13**, e12217 (2021).
- Samaroo, A. et al. Diminished learning over repeated exposures (LORE) in preclinical Alzheimer's disease. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **12**, e12132 (2020).
- Papp, K. V. et al. Early detection of amyloid-related changes in memory among cognitively unimpaired older adults with daily digital testing. *Ann. Neurol.* **95**, 507–517 (2024).
- Lim, Y. Y. et al. Association of deficits in short-term learning and A β and hippocampal volume in cognitively normal adults. *Neurology* **95**, e2577–e2585 (2020).
- Wilks, H. et al. Sharper in the morning: Cognitive time of day effects revealed with high-frequency smartphone testing. *J. Clin. Exp. Neuropsychol.* **43**, 825–837 (2021).
- Muurling, M. et al. Augmented reality versus standard tests to assess cognition and function in early Alzheimer's disease. *Npj Digit. Med.* **6**, 234 (2023).
- van den Berg, R. L. et al. Digital remote assessment of speech acoustics in cognitively unimpaired adults: feasibility, reliability and associations with amyloid pathology. *Alzheimers Res. Ther.* **16**, 176 (2024).
- Jutten, R. J. et al. Detecting early cognitive deficits in preclinical Alzheimer's disease using a remote digital multi-day learning paradigm. *Npj Digit. Med.* **8**, 24 (2025).
- Berron, D. et al. Feasibility of digital memory assessments in an unsupervised and remote study setting. *Front. Digit. Health* **4**, 892997 (2022).
- Lancaster, C. et al. Evaluating the feasibility of frequent cognitive assessment using the Mezurio smartphone app: observational and interview study in adults with elevated dementia risk. *JMIR MHealth UHealth* **8**, e16142 (2020).
- Muurling, M. et al. Feasibility and usability of remote monitoring in Alzheimer's disease. *Digit. Health* **10**, 20552076241238132 (2024).
- Rentz, D. M. et al. The feasibility of at-home iPad cognitive testing for use in clinical trials. *J. Prev. Alzheimers Dis.* **3**, 8–12 (2016).
- Weizenbaum, E. L. et al. Capturing learning curves with the multiday Boston remote assessment of neurocognitive health (BRANCH): Feasibility, reliability, and validity. *Neuropsychology* **38**, 198–210 (2024).
- Moore, R. C. et al. Feasibility and validity of ecological momentary cognitive testing among older adults with mild cognitive impairment. *Front. Digit. Health* **4**, 946685 (2022).
- McLeod, D. R., Griffiths, R. R., Bigelow, G. E. & Yingling, J. An automated version of the digit symbol substitution test (DSST). *Behav. Res. Methods Instrum.* **14**, 463–466 (1982).
- Rentz, D. M. et al. Association of digital clock drawing with PET amyloid and tau pathology in normal older adults. *Neurology* **96**, e1844–e1854 (2021).
- Falletti, M. G., Maruff, P., Collie, A. & Darby, D. G. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J. Clin. Exp. Neuropsychol.* **28**, 1095–1112 (2006).
- Fray, P. J., Robbins, T. W. & Sahakian, B. J. Neuropsychiatric applications of CANTAB. *Int. J. Geriatr. Psychiatry* **11**, 329–336 (1996).
- Spiers, H. J., Coutrot, A. & Hornberger, M. Explaining world-wide variation in navigation ability from millions of people: citizen science project sea hero quest. *Top. Cogn. Sci.* **15**, 120–138 (2023).
- Robin, J., Xu, M., Kaufman, L. D. & Simpson, W. Using digital speech assessments to detect early signs of cognitive impairment. *Front. Digit. Health* **3**, 749758 (2021).
- Gregory, S. et al. Remote data collection speech analysis in people at risk for Alzheimer's disease dementia: usability and acceptability results. *Front. Dement.* **2**, 1271156 (2023).

30. Muurling, M. et al. Remote monitoring technologies in Alzheimer's disease: design of the RADAR-AD study. *Alzheimers Res. Ther.* **13**, 89 (2021).
31. Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *Npj Digit. Med.* **2**, 9 (2019).
32. Popp, Z. et al. Shifting from active to passive monitoring of Alzheimer disease: the state of the research. *J. Am. Heart Assoc.* **13**, e031247 (2024).
33. Wu, J., Chen, N., Xia, H., Wang, Z. & Li, Y. Design and application of a game-based WeChat mini-program for screening cognitive impairments in Chinese older adults. *Alzheimers Dement.* **19**, e077166 (2023).
34. Bardram, J. E. Remote assessment in healthcare—technologies, methods, benefits, and challenges. *PLoS ONE* **18**, e0283945 (2023).
35. Weizenbaum, E. L. et al. Using a digital assessment of multi-day learning curves to detect preclinical AD. *Alzheimers Dement.* **18**, e062194 (2022).
36. Cerino, E. S. et al. Variability in cognitive performance on mobile devices is sensitive to mild cognitive impairment: results from the Einstein aging study. *Front. Digit. Health* **3**, 758031 (2021).
37. Nicosia, J. 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. Neuropsychol. Soc.* **29**, 459–471 (2023).
38. Thompson, L. I. 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.* **14**, e12283 (2022).
39. Sliwinski, M. J. Measurement-burst designs for social health research. *Soc. Personal. Psychol. Compass* **2**, 245–261 (2008).
40. Stawski, R. S., MacDonald, S. W. S. & Sliwinski, M. J. In *The Encyclopedia of Adulthood and Aging* (ed. Whitbourne, S. K.) (Wiley, 2015).
41. Aschenbrenner, A. J. & Jackson, J. J. High-frequency assessment of mood, personality, and cognition in healthy younger, healthy older and adults with cognitive impairment. *Aging Neuropsychol. Cogn.* **31**, 914–931 (2024).
42. Jutten, R. J. et al. A Neuropsychological perspective on defining cognitive impairment in the clinical study of Alzheimer's disease: towards a more continuous approach. *J. Alzheimers Dis.* **86**, 511–524 (2022).
43. Welhaf, M. S. et al. Naturalistic assessment of reaction time variability in older adults at risk for Alzheimer's disease. *J. Int. Neuropsychol. Soc.* **30**, 428–438 (2024).
44. Berron, D. et al. A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. *Npj Digit. Med.* **7**, 79 (2024).
45. Van Niekerk, J. K., Nielsen, T. A., Pasu, S., Shore, S. & Huppert, F. A. The development of new tests of source memory and a new approach to the testing of equivalence of parallel versions. *Aging Neuropsychol. Cogn.* **11**, 416–427 (2004).
46. Berron, D. et al. Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease. *Alzheimers Dement.* **20**, 4775–4791 (2024).
47. Skirrow, C. et al. Validation of a remote and fully automated story recall task to assess for early cognitive impairment in older adults: longitudinal case-control observational study. *JMIR Aging* **5**, e37090 (2022).
48. Lancaster, C. et al. Gallery home: Smartphone-based assessment of long-term memory in adults at risk of Alzheimer's disease. *J. Clin. Exp. Neuropsychol.* **42**, 329–343 (2020).
49. Öhman, F. et al. Unsupervised mobile app-based cognitive testing in a population-based study of older adults born 1944. *Front. Digit. Health* **4**, 933265 (2022).
50. Schlemmer, M. & Desrichard, O. Is Medical environment detrimental to memory? A test of a white coat effect on older people's memory performance. *Clin. Gerontol.* **41**, 77–81 (2018).
51. Shehab, A. & Abdulle, A. Cognitive and autonomic dysfunction measures in normal controls, white coat and borderline hypertension. *BMC Cardiovasc. Disord.* **11**, 3 (2011).
52. Crutzen, C. & Adam, S. "What if it's not just an item of clothing?" – a narrative review and synthesis of the white coat in the context of aged care. *Psychol. Belg.* **62**, 62–74 (2022).
53. Hauser, D., Paolacci, G. & Chandler, J. in *Handbook of Research Methods in Consumer Psychology* (eds. Kardes, F. R., Herr, P. M. & Schwarz, N.) (Taylor & Francis Group, 2019).
54. Nicosia, J. et al. Bridging the technological divide: stigmas and challenges with technology in digital brain health studies of older adults. *Front. Digit. Health* **4**, 880055 (2022).
55. International Telecommunication Union. Measuring digital development: facts and figures 2023. (2023).
56. Hou, Y. et al. Ageing as a risk factor for neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 565–581 (2019).
57. Wang, A. -Y. et al. Socioeconomic status and risks of cognitive impairment and dementia: a systematic review and meta-analysis of 39 prospective studies. *J. Prev. Alzheimers Dis.* **10**, 83–94 (2023).
58. Siddiqua, A., Karim, A. & Gani, A. Big data storage technologies: a survey. *Front. Inf. Technol. Electron. Eng.* **18**, 1040–1070 (2017).
59. Filkins, B. L. et al. Privacy and security in the era of digital health: what should translational researchers know and do about it? *Am. J. Transl. Res.* **8**, 1560–1580 (2016).
60. Romansky, R. P. & Noninska, I. S. Challenges of the digital age for privacy and personal data protection. *Math. Biosci. Eng.* **17**, 5288–5303 (2020).
61. Goldsack, J. C. et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). *NPJ Digit. Med.* **3**, 55 (2020).
62. Papp, K. V. et al. Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **13**, e12243 (2021).
63. Weiner, M. W. et al. Brain health registry updates: an online longitudinal neuroscience platform. *Alzheimers Dement.* **19**, 4935–4951 (2023).
64. Malzbender, K. et al. Validation, deployment, and real-world implementation of a modular toolbox for Alzheimer's disease detection and dementia risk reduction: the AD-RIDDLE project. *J. Prev. Alzheimers Dis.* **11**, 329–338 (2024).
65. Kaye, J. et al. Using digital tools to advance Alzheimer's drug trials during a pandemic: the EU/US CTAD task force. *J. Prev. Alzheimers Dis.* **8**, 513–519 (2021).
66. Leuzy, A. et al. REAL AD-validation of a realistic screening approach for early Alzheimer's disease. *Alzheimers Dement.* **20**, 8172–8182 (2024).
67. Gregory, S. et al. Remote data collection speech analysis and prediction of the identification of Alzheimer's disease biomarkers in people at risk for Alzheimer's disease dementia: the speech on the phone assessment (SPeAk) prospective observational study protocol. *BMJ Open* **12**, e052250 (2022).
68. König, A. et al. Screening over speech in unselected populations for clinical trials in AD (PROSPECT-AD): study design and protocol. *J. Prev. Alzheimers Dis.* **10**, 314–321 (2023).
69. Stricker, N. H. et al. Stricker learning span criterion validity: a remote self-administered multi-device compatible digital word list memory measure shows similar ability to differentiate amyloid and tau PET-defined biomarker groups as in-person auditory verbal learning test. *J. Int. Neuropsychol. Soc.* **30**, 138–151 (2024).
70. Boots, E. A. et al. Continuous associations between remote self-administered cognitive measures and imaging biomarkers of Alzheimer's disease. *J. Prev. Alzheimers Dis.* **11**, 1467–1479 (2024).
71. Thompson, L. I. et al. Remote and in-clinic digital cognitive screening tools outperform the MoCA to distinguish cerebral amyloid status among cognitively healthy older adults. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **15**, e12500 (2023).

72. Toniolo, S. et al. Relationship of plasma biomarkers to digital cognitive tests in Alzheimer's disease. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **16**, e12590 (2024).
73. Shiffman, S., Stone, A. A. & Hufford, M. R. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* **4**, 1–32 (2008).
74. Trull, T. J. & Ebner-Priemer, U. Ambulatory assessment. *Annu. Rev. Clin. Psychol.* **9**, 151–176 (2013).
75. Jutten, R. J. et al. Monthly at-home computerized cognitive testing to detect diminished practice effects in preclinical Alzheimer's disease. *Front. Aging Neurosci.* **13**, 800126 (2022).
76. Fristed, E. et al. A remote speech-based AI system to screen for early Alzheimer's disease via smartphones. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **14**, e12366 (2022).
77. Weiner, M. W. et al. Increasing participant diversity in AD research: plans for digital screening, blood testing, and a community-engaged approach in the Alzheimer's disease neuroimaging initiative 4. *Alzheimers Dement.* **19**, 307–317 (2023).
78. Saunders, S., Haider, F., Ritchie, C. W., Muniz Terrera, G. & Luz, S. Longitudinal observational cohort study: speech for Intelligent cognition change tracking and DEtection of Alzheimer's disease (SIDE-AD). *BMJ Open* **14**, e082388 (2024).
79. Alty, J. et al. Development of a smartphone screening test for preclinical Alzheimer's disease and validation across the dementia continuum. *BMC Neurol.* **24**, 127 (2024).
80. Miller, M. J. The ADNI4 digital study: a novel approach to recruitment, screening, and assessment of participants for AD clinical research. *Alzheimers Dement.* **20**, 7232–7247 (2024).
81. ter Huurne, D. et al. User experience of a (semi-) automated cognitive phone-based assessment within a memory clinic population. *Arch. Clin. Neuropsychol.* **40**, 319–329 (2025).
82. Sliwinski, M. J. et al. Reliability and validity of ambulatory cognitive assessments. *Assessment* **25**, 14–30 (2016).
83. Thompson, L. I. et al. Psychometric and adherence considerations for high-frequency, smartphone-based cognitive screening protocols in older adults. *J. Int. Neuropsychol. Soc.* **30**, 785–793 (2024).
84. Tröger, J. et al. Validation of the remote automated ki:e speech biomarker for cognition in mild cognitive impairment: verification and validation following DiME V3 framework. *Digit. Biomark.* **6**, 107–116 (2022).
85. Stricker, N. H. et al. A novel computer adaptive word list memory test optimized for remote assessment: psychometric properties and associations with neurodegenerative biomarkers in older women without dementia. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **14**, e12299 (2022).
86. Papp, K. V., Rentz, D. M., Orlovsky, I., Sperling, R. A. & Mormino, E. C. Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: the PACC5. *Alzheimers Dement. Transl. Res. Clin. Interv.* **3**, 668–677 (2017).
87. Donohue, M. C. et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol.* **71**, 961 (2014).
88. Pratap, A. et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *Npj Digit. Med.* **3**, 21 (2020).
89. Li, S. X. et al. Recruitment and retention in remote research: learnings from a large, decentralized real-world study. *JMIR Form. Res.* **6**, e40765 (2022).
90. Bilder, R. M. & Reise, S. P. Neuropsychological tests of the future: how do we get there from here?. *Clin. Neuropsychol.* **33**, 220–245 (2019).
91. Cid, R. E. C. & Loewenstein, D. A. Salient cognitive paradigms to assess preclinical Alzheimer's disease. *Neurotherapeutics* **19**, 89–98 (2022).
92. Singh, S. & Germine, L. Technology meets tradition: a hybrid model for implementing digital tools in neuropsychology. *Int. Rev. Psychiatry* **33**, 382–393 (2021).
93. Mank, A. et al. Identifying relevant outcomes in the progression of Alzheimer's disease; what do patients and care partners want to know about prognosis? *Alzheimers Dement. Transl. Res. Clin. Interv.* **7**, e12189 (2021).
94. Lott, S. A. et al. Digital health technologies for Alzheimer's disease and related dementias: initial results from a landscape analysis and community collaborative effort. *J. Prev. Alzheimers Dis.* **11**, 1480–1489 (2024).
95. Peters, M. D. J. et al. Guidance for conducting systematic scoping reviews. *Int. J. Evid. Based Healthc.* **13**, 141–146 (2015).
96. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan — a web and mobile app for systematic reviews. *Syst. Rev.* **5**, 210 (2016).
97. Tricco, A. C. et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med.* **169**, 467–473 (2018).
98. Maruff, P. et al. Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychol.* **1**, 30 (2013).

Acknowledgements

We thank Merle Hinz and Sarah Kriener for their help checking duplicate reports and labeling excluded reports. This work was supported by the NIH/NIA (1R01AG084017-01A1) and the EU Joint Programme—Neurodegenerative Disease Research (JPND) expert group Remote Digital Assessment and Monitoring for Early Alzheimer's Disease (REMOTE-AD), which is supported by the German Federal Ministry of Education and Research (BMBF; 01ED2401) under the aegis of JPND (www.jpnd.eu).

Author contributions

SEP: Literature search, report screening, writing; FÖ: Literature search, report screening, writing; JH: Writing; AK: Writing; KVP: Writing; MS: Conceptualization, writing; DB: Conceptualization, writing. All authors have read and approved the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

SEP and FÖ declare no competing interests. JH is a paid consultant for Eisai, AlzPath, Prothena. AK is an employee of ki:elements. KVP has served as a paid consultant for Novoic, Prothena, and Biogen and is on the advisory board for Cogstate. MS has served on advisory boards for Roche and Novo Nordisk, received speaker honoraria from Bioarctic, Eisai, Genentech, Lilly, Novo Nordisk and Roche and receives research support (to the institution) from Alzpath, Bioarctic, Novo Nordisk and Roche (outside scope of submitted work); he is a co-founder and shareholder of Centile Bioscience and serves as Associate Editor with *Alzheimer's Research & Therapy*. DB is co-founder and shareholder of neotiv GmbH.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-025-01583-5>.

Correspondence and requests for materials should be addressed to Sarah E. Polk or David Berron.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025