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

Polk, S. E.[1,*], Öhman, F.[2,3,4], Hassenstab, J.[5,6], König, A.[7,8,9], Papp, K. V.[10], Schöll, M.[2,3,4,11,†], Berron, D.[1,12,13,*,†]

1. Clinical Cognitive Neuroscience, German Center for Neurodegenerative Diseases (DZNE), Magdeburg, DE
2. Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, SE
3. Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, SE
4. Region Västra Götaland, Sahlgrenska University Hospital, Department of Neuropsychiatry, Gothenburg, SE
5. Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA
6. Department of Neurology, Washington University in St. Louis School of Medicine, St. Louis, MO, USA
7. ki:elements UG, Saarbrücken, DE
8. Cognition Behaviour Technology (CoBTek) Lab, University Côte d'Azur, Nice, FR
9. Université Côte d'Azur, Centre Hospitalier et Universitaire, Clinique Gériatrique du Cerveau et du Mouvement, Centre Mémoire de Ressources et de Recherche, Nice, FR
10. Mass General Brigham, Department of Neurology, Harvard Medical School, Boston, MA, USA
11. Dementia Research Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, UK
12. Center for Behavioral Brain Sciences, Otto-von-Guericke University Magdeburg, Magdeburg, DE
13. Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Lund, SE

*Corresponding authors: Sarah E. Polk, Clinical Cognitive Neuroscience, German Center for Neurodegenerative Diseases (DZNE), Magdeburg, DE; E-mail: sarah.polk@dzne.de; David Berron, Clinical Cognitive Neuroscience, German Center for Neurodegenerative Diseases (DZNE), Magdeburg, DE; E-mail: david.berron@dzne.de.
[†]MS and DB contributed equally to this manuscript

Abstract

Subtle cognitive changes in preclinical Alzheimer's disease (AD) are difficult to detect using traditional pen-and-paper neuropsychological assessments. Remote and unsupervised digital assessments can improve scalability, measurement reliability, and ecological validity, enabling the detection and monitoring of subtle cognitive change. Here, we evaluate such tools deployed in preclinical AD samples, defined as cognitively unimpaired individuals with abnormal levels of amyloid-β (Aβ), or Aβ and tau.

In this scoping review, we screened 1,680 unique reports for studies using remote and unsupervised cognitive assessment tools in preclinical AD samples; 23 tools were found. We describe each tool's usability, validity, and reported metrics of reliability. Construct and criterion validity according to associations with established neuropsychological assessments and measures of Aβ and tau are reported.

With this review, we aim to present a necessary update to a rapidly evolving field, following a previous review by Öhman and colleagues (2021; *Alzheimers Dement. Diagn. Assess. Dis. Monit*) and addressing the open questions of feasibility and reliability of remote testing in the target population. We discuss future directions for using remote and unsupervised digital cognitive assessments in preclinical AD and how such tools may be used for longitudinal monitoring of cognitive function, scalable case finding, and individualized prognostics in both clinical trials and healthcare contexts.

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.

Declaration of Competing Interests: SEP and FÖ have nothing to declare. 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.

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).

Introduction

Subtle cognitive changes that emerge in the preclinical stage of Alzheimer's disease (AD) are difficult to detect using conventional pen-and-paper neuropsychological assessments[1,2]. This presents a challenge, as detecting and monitoring the earliest cognitive changes caused by AD is becoming paramount, particularly as clinical trials aim to screen and enroll participants at risk for cognitive decline before the clinical symptoms of Alzheimer's disease appear[3]. 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[a] and TRAILBLAZER-ALZ 3[b]. Once early treatments become available, tools that can reliably detect and predict subtle cognitive decline due to AD on an individual level will be essential[4]. 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 pathological build-up of amyloid-β (Aβ) plaques and neurofibrillary tau tangles that characterize AD[5–7], before clinical symptoms manifest (i.e., preclinical AD).

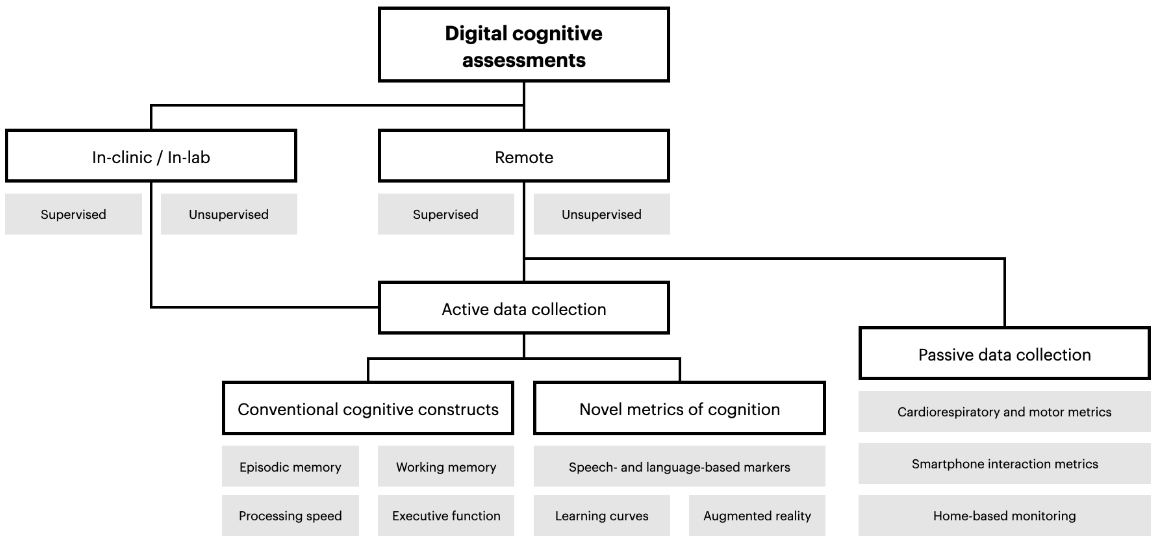
Digital cognitive assessments: A brief taxonomy

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)[7]. A wide range of existing pen-and-paper assessments have been digitized[8,9], and a growing number of cognitive tests are being developed specifically for digital administration (e.g., refs. [10–12]). In addition to tests that

[a] https://clinicaltrials.gov/study/NCT04468659
[b] https://clinicaltrials.gov/study/NCT05026866

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. [13,14]) or multi-modal assessments (i.e., assessment of both active and passive markers of cognition simultaneously[15]). 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[16,17]. See Figure 1 for an overview of the clinical applications of digital cognitive assessments.

Figure 1.
A non-exhaustive taxonomy of the clinical applications of digital cognitive assessments. The current review focuses on those assessments that are remotely deployed without supervision, and which use active data collection to quantify cognitive function.



Benefits of and challenges associated with remote and unsupervised testing

The digitization of cognitive assessments enables the remote and unsupervised collection of cognitive data using participants' own devices, which comes with a number of benefits as well as challenges. For example, 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. [12,18]) as well as web-based cognitive tests (e.g., TestMyBrain[c]). Additionally, the frequency of within-person testing can be increased. Individuals can feasibly self-administer remote assessments once a month for a year[19], on a daily basis[20,21], or even multiple times a day[22–24]. Measurement burst designs can also be implemented (e.g., a week of daily assessments every six months for several years)[25,26]. 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[27–29]. 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[21,30,31], recall after several hours or days[32–34], or the effects of time of day[35].

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"[36,37]. Performance on remote assessments may be more reflective of individuals' everyday cognitive function, as it may be less associated with increased anxiety during testing[38]. On the other hand, unproctored testing introduces challenges related to environmental distractions, low effort, and malingering, as well as data fidelity. Building attention checks into tasks and asking explicitly if participants were distracted while completing

[c] https://www.testmybrain.org/index.html

the tasks may limit these effects. Researchers may also consider integrating reliable participant authentication processes.

Another challenge relates to the level of digital literacy necessary for individuals to complete digital testing without assistance, which can vary depending on the target population[39]. 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 Union[40], 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 the rest of the world, as well as in rural areas and low-income countries[40]. These statistics should be borne in mind when planning remote data collection.

Finally, the underlying infrastructure of 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 technologies[41] 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[42,43]. Researchers should understand the risks associated with data transfer and storage to make informed decisions that comply with relevant regulatory frameworks.

The current review

With these benefits and challenges in mind, as well as the current need for sensitive cognitive testing in preclinical AD, this 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 Aβ, or Aβ and tau. We aim to present a necessary update of this rapidly evolving field since the publication of our previous review[7], 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+ Framework[44], which was developed as a common framework for digital health technologies measuring a wide range of metrics[d]), 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 Aβ (and tau) burden, whether they correlate with continuous measures of Aβ 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[16,17], 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

[d] https://datacc.dimesociety.org/v3/

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.

Methods

Search strategy and selection criteria

An initial literature search was performed on September 12th, 2023 using PubMed, Web of Science, and APA PsycINFO using terms including digital, remote, unsupervised, smartphone, cognition, Alzheimer's, and dementia. Publications already known to the authors were also included. Identical literature searches were repeated on March 8th and September 9th, 2024. All records were uploaded onto Rayyan, a web-based literature screening tool[45]. Duplicates were automatically identified, then manually checked and excluded, as were records in languages other than English. The remaining 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., measures of Aβ 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.

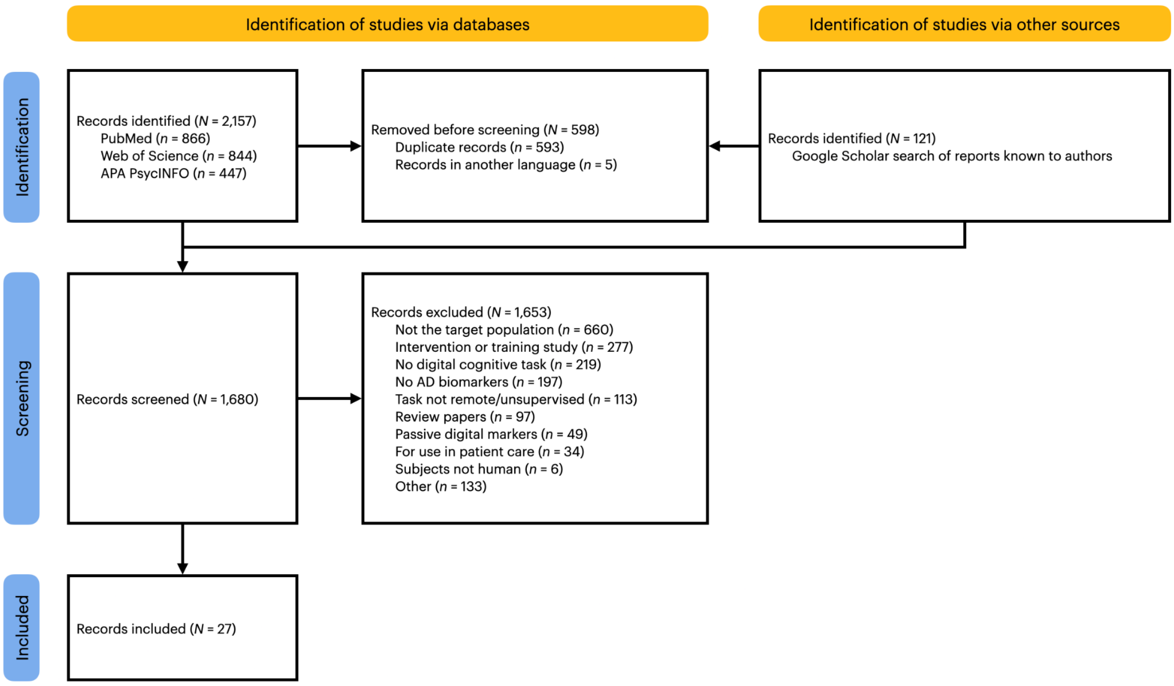
If certain feasibility or psychometric information about the tools was not reported in the included articles, this was gathered from other papers using the same tools, sometimes in separate samples, found using Google Scholar. Additionally, if unstandardized estimates were reported, these estimates were standardized using standard errors or 95% confidence intervals and *N*s to be able to compare effect sizes; these calculations were not done with the original data and are considered approximations.

Results

Literature screening

A total of 2,278 records were found, 598 of which were excluded as duplicates or records in languages other than English. The remaining 1,680 were screened and 27 relevant reports were found (see Figure 2 for details), two of which were previously included[7]. These and the 23 tools described in these records are listed in Table 1.

Figure 2.
PRISMA flow diagram detailing the screening of records.



Note. Some articles were labeled with multiple exclusion reasons, and thus the sum of exclusion reasons exceeds the number of records excluded.

Both conventional and novel metrics of cognition can be captured remotely

Seventeen of the included tools capture the conventional cognitive constructs: Altoida[46], Ambulatory Research in Cognition (ARC)[23,35], Boston Remote Assessment for Neurocognitive Health (BRANCH)[47], Cambridge Neuropsychological Test Automated Battery (CANTAB)[48,49], Cogstate Brief Battery (CBB)[48,50,51], cCOG[49], Cognitron[52], the ki:elements speech biomarker for cognition (ki:e SB-C)[53,54], Mayo Test Drive (MTD)[55,56], MemTrax Memory Test[48], Mezurio[15], Mobile Monitoring of Cognitive Change (M2C2)[57], neotiv[32,49,52,58], NeuroVocalix[49], NIH Mobile Toolbox[48], Oxford Cognitive Testing Portal (OCTAL)[59], and the ReVeRe word list recall test (RWLRT)[48]. 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), two (ARC and M2C2) use ecological momentary assessment-based paradigms[60], one (BRANCH) is used to quantify learning of repeated stimuli over days to months (i.e., learning curves)[30], and four (ki:e SB-C, Mezurio, NeuroVocalix, RWLRT) evaluate cognitive function using speech-based metrics. Two other tools are used to quantify learning curves: Computerized Cognitive Composite (C3) Face Name Associative Memory Exam (FNAME)[19,61], and Online Repeatable Cognitive Assessment—Language Learning Test (ORCA-LLT) [31]. Finally, an additional four tools use speech-based metrics: Novoic[62,63], the Speech for Intelligent cognition change tracking and DEtection of Alzheimer's Disease (SIDE-AD) online platform[64], Winterlight Assessment (WLA)[65], and TapTalk[66], 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, Cogstate) or remote digital tools[7], but had not yet been validated for remote use in preclinical AD (except BRANCH and ORCA-LLT[19,31]).

Tools that have been validated for use in preclinical AD

Feasibility of remote and unsupervised data collection

These studies were generally highly feasible as shown by rates of consent into studies, enrollment, adherence, compliance, and user experience reports. Three studies, all of which 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%[19], and a week-long measurement burst design[25,26] with four daily measurements and had a consent rate of 87%[39]. Finally, a longitudinal paradigm with biweekly assessments for 12 months also reported an 86% consent rate[32]. However, the latter also reported that 24% of those who consented did not enroll in the app[32].

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[23,30,35,57,62,65], from 63% to 94% for an eight-week study[67], and from 75% to 78% in year-long studies[31,61]. Compliance, defined here as the completion of measurements as intended (e.g., acceptable data quality, successful attention checks), was generally excellent, with only 2–3% of data being unusable in both cross-sectional and longitudinal designs measuring conventional cognitive metrics as well as learning curves[47,59,61,65]. 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 reasons[46]. Similarly, preliminary feasibility results from ADNI4 showed that only 54% of eligible individuals completed speech-based tasks[68].

Regarding user experience, in a cross-sectional paradigm, 16% of participants reported they had technical difficulties[47], and in a week-long speech-based paradigm, 15% reported difficulties[69], while in a year-long design, 30% of participants reached out for technical support[19]. Tools were generally straightforward and enjoyable to use[23,30,32,35,47,57,65,69,70], with some participants indicating that they would repeat or continue such studies[57].

Reliability of remote and unsupervised data collection

We also evaluated the reliability of each tool, insofar as this was reported (see Table 1). Between-person reliability (i.e., variability across individuals, calculated with intraclass correlations according to ref. [71]) was good for ARC and across multiple M2C2 assessments[23,35,72]. Parallel-forms reliability (i.e., reliability across alternate versions of the same task) was good for C3 FNAME[73], as well as for Novoic on average across versions[62]; no other studies reported parallel-forms reliability. Test-retest reliability (i.e., precision of measurement within an individual) calculated as intraclass or Pearson's correlations was moderate to excellent for most tasks[23,32,47,74–76], poor for Altoida[46], and varied widely across WLA metrics[65], though most reached ICCs ≥ .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).

Those tools that reported associations with measures of global cognition (e.g., Preclinical Alzheimer's Cognitive Composite [PACC][77,78] or similar composites) found correlations coefficients (rs and ρs) between |.53| and |.70| and a β estimate of .26[23,32,46,47]. Two studies found associations between remote tasks and traditional neuropsychological tests measuring the same cognitive constructs (rs = |.59|–|.61|; βs = |.22|–|.44|)[72,76]. Another study used speech-based metrics to predict PACC5 and found that predicted and measured PACC5 scores correlated (r = .74)[62]. Finally, some tools reported associations of baseline performance or learning curves on remote tasks with retrospective, concurrent, or subsequent change in PACC scores (*r*s = .54–.69, *β* = 0.27–0.59)[30,31,61].

Regarding associations with biomarkers of Aβ and/or tau pathology, a number of studies reported differences in performance between Aβ– and Aβ+ groups[19,30,31,35,46,65]. Most group differences had medium effect sizes (Cohen's *d*s = 0.49–0.66, *β*s = |0.30|–|0.40|), with Aβ+ groups performing worse than Aβ– groups. One study found a small effect of time of day (Cohen's *d* = 0.19), with individuals with elevated p-tau181/Aβ42 levels performing worse in the evening[35], while another found a very large effect of Aβ on learning rates across a year (Cohen's *d* = 2.22), with a Aβ+ group showing dramatically slowed learning[31]. A number of studies reported associations between performance on remote tasks and continuous measures of Aβ and/or tau[19,23,31,32,47,56,59,61], with correlation coefficients ranging from |.11| to |.34| and *β*s ranging from |0.23| to |0.38|. Finally, some studies reported that remote task performance could distinguish between Aβ– and Aβ+ individuals[46,55,57,62] or Aβ/tau– and Aβ/tau+ individuals[55], with areas under the curve between .63 and .83.

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.

Table 1.
Remote and unsupervised digital cognitive assessments for use in preclinical AD samples in alphabetical order. For samples including groups other than preclinical AD, effect sizes are reported for the findings pertaining to the preclinical AD sub-sample.

					Analytical and clinical validity according to associations with:	
Tool	**Report included**	***N* included in analyses or ongoing study**	**Cognitive domain (task)**	**Reliability**	**Neuropsychological tests**	**Aβ/tau**
Altoida	Muurling et al., 2023[46]	**56** (29 PET or CSF Aβ–/27 Aβ+) + 65[b]	**Episodic memory** (Augmented reality tasks) **High-dimensional multi-modal data**	Test-retest reliability: ICC = .48	Corr. with a cognitive composite (ρ = .56)	Lower scores in Aβ+ than Aβ– ($\beta \approx -.40$)[g]; distin. Aβ– vs. Aβ+ (AUC = .76)
ARC	Wilks et al., 2021[35]	**169** (81 CSF pTau:Aβ42–/32 pTau:Aβ42+)	**Episodic memory** (Prices) **Processing speed** (Symbols) **Working memory** (Grids) **EMA-based**	Between-person reliability: ICCs > .80		Sundowning in CSF p-tau:Aβ42+ group (Cohen's d = .19)
	Nicosia et al., 2023[23]	**290** (212 with PET, 146 with CSF)		Between-person reliability: ICCs > .81; test-retest reliability: ICCs > .85	Corr. with a cognitive composite (r = –.53)[h]	Corr. with Aβ (r = .26) and tau PET (.11); CSF Aβ42 (–.23), t-tau (.28), p-tau181 (.25)[h]
BRANCH	Papp et al., 2021[47]	**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 = .81	Corr. with PACC (r = .62)	Corr. with cortical Aβ (r = –.21) and entorhinal tau PET (–.18)
	Papp et al., 2024[30]	**164** (128 PET Aβ–/36 Aβ+)		Test-retest reliability: ICC = .94[74]	Corr. with subsequent decline on PACC (r = .54)	Reduced learning rates in Aβ+ (Cohen's d = 0.49)
CANTAB	Weiner, Aaronson, et al., 2023[48]	**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) **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**	
	Malzbender et al., 2024[49]	**AD-RIDDLE**			**Awaiting validation in preclinical AD**	
CBB	Langford et al., 2020[50]	**4,486** (3,163 PET Aβ–/1,323 Aβ+)		Test-retest reliability: ICC = .90–.95[79]		Distin. Aβ– vs. Aβ+ (AUCs = .60–.73)
	Kaye et al., 2021[51]	**TRC-PAD**			**Awaiting validation in preclinical AD**	

	Weiner, Aaronson, et al., 2023[48]	**Brain Health Registry**	**Episodic memory** (One Card Learning) **Working memory** (One-Back) **Executive function** (Identification) **Psychomotor** (Detection)		**Awaiting validation in preclinical AD**	
cCOG	Malzbender et al., 2024[49]	**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[52]	**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**	
C3 FNAME	Samaroo et al., 2020[19,a]	**94** (69 PET Aβ–/25 Aβ+; 84 with FTP)	**Episodic memory** (Face Name Associative Memory Exam) **Learning curves**	Parallel-forms reliability: Cronbach's $\alpha > .85$[73]		Reduced learning curves in Aβ+ (Cohen's d = 0.66), corr. with tau (r = –.22)
	Jutten et al., 2022[61]	**114** (81 PET Aβ–/33 Aβ+; 105 with FTP)			Corr. with concurrent PACC5 decline (r = .69)	Corr. with global Aβ (r = –.20), entorhinal tau (–.38), and inf.-temp. tau PET (–.23)
ki:e SB-C	Gregory et al., 2022[53]	**SPeAk**	**Structured speech-based tasks**	Test-retest reliability: rs = .57–.72[75]	**Awaiting validation in preclinical AD**	
	König et al., 2023[54]	**PROSPECT-AD**			**Awaiting validation in preclinical AD**	
MemTrax Memory Test	Weiner, Aaronson, et al., 2023[48]	**Brain Health Registry**	**Episodic memory**		**Awaiting validation in preclinical AD**	
Mezurio	Muurling et al., 2021[15]	**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[55]	**353** (228 PET Aβ–/125 Aβ+; 250 with FTP)	**Episodic memory** (Stricker Learning Span) **Processing speed** (Symbols)	Test-retest reliability: ICCs > .71[76]	Corr. with in-person NPTs of same construct (rs = |.26|–|.51|)[76]	Distin. Aβ– vs. Aβ+ (AUCs = .63–.77), Aβ–/tau– vs. Aβ+/tau+ (.67–.83)

	Boots et al., 2024[56]	**684** (670 with PiB, 667 with FTP)			Corr. with PACC (r = .68); corr. with in-person NPTs of same construct (rs = |.59|–|.61|)	Corr. with Aβ (ρ = –0.24), entorhinal tau (–0.23), and global tau PET (–0.21)
M2C2	Thompson et al., 2023[57]	**69** (44 PET Aβ–/25 Aβ+)	**Episodic memory** (Prices) **Processing speed** (Symbol Match) **Working memory** (Color Shapes) **EMA-based**	Between-person reliability: ICCs = .25–.97[72,e]	Corr. with in-person NPTs of same construct (βs = |.22|–|.44|)[72]	Distin. Aβ– vs. Aβ+ (AUCs = .73–.77)
neotiv	Berron, Olsson et al., 2024[32]	**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 = .65–.83	Corr. with PACC (r = .62–.70); pred. concurrent PACC decline (βs ≈ 0.41–0.59)[i]	Pred. Aβ in precuneus (βs ≈ .23–.38) and tau in the MTL (.20–.40)[i]
	Berron, Olsson et al., 2024[32]	**BIOFINDER-2**		**Awaiting further validation in preclinical AD**		
	Berron, Glanz et al., 2024[58]	**DELCODE, WRAP**		**Awaiting further validation in preclinical AD**		
	Malzbender et al., 2024[49]	**AD-RIDDLE**		**Awaiting further validation in preclinical AD**		
	Leuzy et al., 2024[52]	**REAL AD**		**Awaiting further validation in preclinical AD**		
NeuroVocalix	Malzbender et al., 2024[49]	**AD-RIDDLE**	**Speech-based episodic memory**	**Awaiting validation in preclinical AD**		
NIH Mobile Toolbox	Weiner, Aaronson, et al., 2023[48]	**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**		
Novoic	Fristed et al., 2022[62]	**115** (59 PET or CSF Aβ–/56 Aβ+)	**Speech-based storytelling**	Parallel-forms reliability: ρ = .39–.85[69,f]	Predicted PACC corr. with measured PACC (r = .74)	Distin. Aβ– vs. Aβ+ (AUC = .74)
	Weiner, Veitch, et al., 2023[63]	**ADNI4**		**Awaiting further validation in preclinical AD**		
OCTAL	Toniolo et al., 2024[59]	**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 = .34); pred. plasma p-tau181[j]
ORCA-LLT	Lim et al., 2020[31,a]	**80** (42 PET Aβ–/38 Aβ+)	**Learning curves**		Pred. baseline EM (β = 0.26); pred. retrospective EM change (β = 0.27)	Reduced learning curves in Aβ+ (Cohen's d = 2.22); pred. Aβ PET (β = –0.23)

RWLRT	Weiner, Aaronson, et al., 2023[48]	**Brain Health Registry**	**Speech-based episodic memory**		**Awaiting validation in preclinical AD**	
SIDE-AD Platform	Saunders et al., 2024[64]	**SIDE-AD**	**Unstructured speech-based tasks**		**Awaiting validation in preclinical AD**	
TapTalk	Alty et al., 2024[66]	**Validation studies of TapTalk**	**Speech- and motor-based task**		**Awaiting validation in preclinical AD**	
WLA	van den Berg et al., 2024[65]	**50** (27 PET Aβ–/23 Aβ+)	**Structured and unstructured speech tasks**	Test-retest reliability: ICCs = −.06 to .97[g]		More pauses in Aβ+ than Aβ– (βs ≈ .30–.34)[i]

Note. 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 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.

[a] Included in ref. [7]

[b] n = 65 did not complete remote testing.

[c] n = 352 were included as an online sample with no biomarkers.

[d] Only Paired Associates Learning was announced to be used in association with the Brain Health Registry.

[e] Between-person reliability was greatly improved for the M2C2 when aggregating multiple measurements (≥ 0.76).

[f] Novoic tested 153 correlations across test versions, and average ρs reached .73.

[g] WLA tested 186 ICCs; reliability reached ICCs of ≥ .50 across two or more assessments.

[h] Higher ARC score indicates worse performance, thus the correlations reported here are in the expected direction.

[i] Unstandardized estimates originally reported were standardized using the standard errors or 95% confidence intervals and Ns reported.

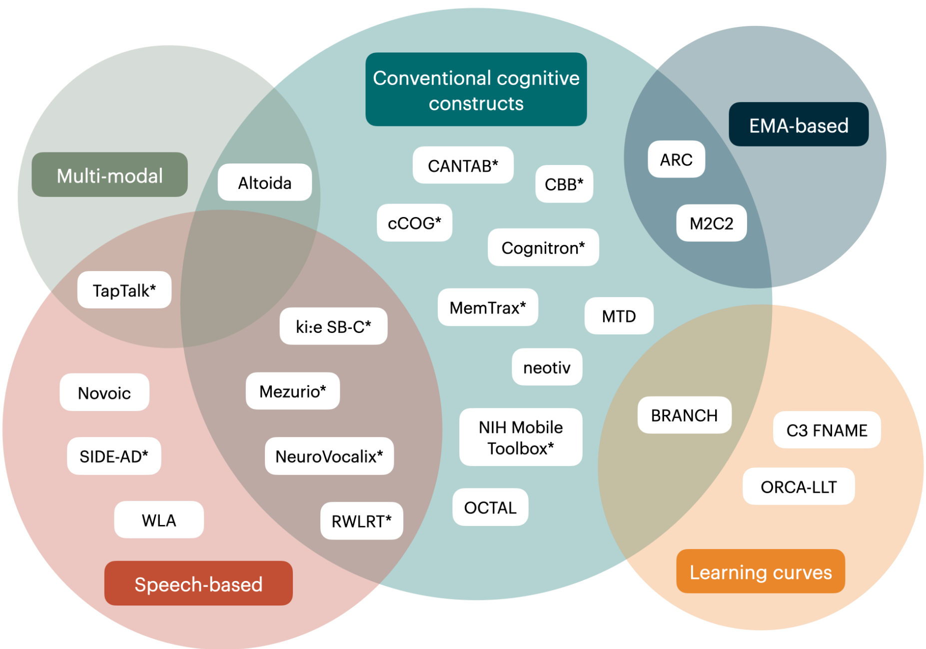
[j] Only p-values were reported for the models including individual metrics. The best model to predict p-tau181 combined remote and in-person metrics, as well as Aβ42/40, adjusted R^2 = .50.

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 Aβ, 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.

Figure 3.

Venn diagram of the tools included in the current review based on the type of cognitive metrics they quantify.



Note. 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 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.

The current state of the field

Remote and unsupervised cognitive assessment is feasible with certain limits

At the time our previous review was published[7], 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[19,31]) 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 86–87%, adherence rates from 63–93%, and compliance rates from 97–98%. 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 weeks[80]. Finally, 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.

One potential bottleneck to participant retention may be registration in digital apps after consenting into 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 task[32]. Similarly, of those individuals who met ADNI4 inclusion criteria, only 54% completed speech-based remote tasks[68]. Another factor that may ameliorate participant drop-out is the technical support available; while 15–16% of participants in cross-sectional or week-long studies reported technical difficulties[47,69], 30% of participants in a year-long longitudinal study reached out to the study team for technical support[19]. 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 reasons[46]. 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 app including conventional cognitive tasks had an average problem score of .31, while the augmented reality-based app scored .60 on average[67].

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 WEIRD (Western, Educated, Industrial, Rich, and Democratic) populations. One way of boosting retention, for example, is by having participants engage with the digital tool immediately after consenting into a study[81]. Having technical support available to participants, especially those with lower digital literacy[39], may also alleviate both issues of limited enrollment as well as prohibitive technical difficulties during tasks.

Reliability of remote tests is generally acceptable but metrics should be reported more often

Another open point in the previous review pertained to the reliability of remotely assessed cognitive metrics[7]. 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[23,35,72]; 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 versions[73], while Novoic tasks had moderate to excellent reliability across versions[69]. 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[23,32,47,74,76]. The test-retest reliability for the speech-based WLA varied depending on the metric and the number of assessments across which performance was averaged[65]. For both neotiv and WLA, test-retest reliability improved when aggregating across multiple measurements[32,65]. Altoida was the only tool that reported poor retest reliability[46], which was acceptable for Android users, ICC = .70, but very poor for iOS users, ICC = .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 others.

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 becomes the goal of more and more studies, 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.

Construct validity with established neuropsychological tests may not be the gold standard

The 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. In general, associations were moderate to strong, however no correlations greater than .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 with a Spearman's *ρ* of .57 across at-home and in-clinic[46], as was C3 (composite score across all tasks including FNAME) with an *r* of .70[73], while a remote version of the neotiv Mnemonic Discrimination Task correlated with a similar in-scanner version with an *r* of .66[32]. This suggests that remote and in-clinic assessments, even when using nearly identical measures, may only be correlated to a limited degree.

Additionally, Stricker and colleagues[55] 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 imperfect[82]. 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[82–84]. 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 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., neotiv[32]). Additionally, the target population should again be taken into account here; for example, there may be an element of digital familiarity among older adults[39], 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 Aβ– and Aβ+ groups[55]) may lead to greater precision in measuring behaviors that are meaningful for preclinical AD.

Remote tests are sensitive to cognitive change related to Alzheimer's pathology

Necessarily, all tools included in this review were used to find associations between cognition and markers of Aβ and/or tau pathology. Most studies used PET-characterized Aβ and tau burden[19,23,30–32,46,47,55,57,61,62,65], while some used CSF markers, either as the main markers of interest or as a substitute in the case of missing PET scans[23,35,46,62,65]. Most studies reported known-group validity, comparing Aβ– to Aβ+ groups and finding poorer performance on remote cognitive tests among cognitively healthy individuals with elevated Aβ burden, showing that many of these assessments are sensitive to the subtle cognitive changes in preclinical AD. Associations between continuous measures of Aβ 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 show that remote and unsupervised cognitive assessments can capture 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 Aβ42/40, and cognitive performance on remote digital tasks[59], 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 Aβ42/40 only showed weak correlations. Another study combined performance on a remote task with plasma p-tau217 to predict future decline in PACC score in a mostly cognitively unimpaired sample[32]. 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 invasive procedures such as PET scans, lumbar punctures, and extensive in-person cognitive batteries for individuals with minimal cognitive complaints.

Moving forward with remote and unsupervised digital cognitive assessments

Measuring cognition longitudinally in larger, more diverse samples

Many remote and unsupervised tools are 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 trials[4], will be greatly facilitated. These efforts are especially timely given the recent update to the FDA'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 impairment[3].

Towards individualized medicine

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 real-world health care 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 Aβ burden: two tasks that used conventional cognitive tests, the MDT Stricker Learning Span (verbal learning, which also distinguished between Aβ/tau– and Aβ/tau+)[55] and the M2C2 tasks (the associative memory task performed best)[57], Altoida, which used an augmented reality-based task[46], and Novoic, which employed speech-based tasks[62]. The best performing task achieved an area under the curve of .77 (.83 for Aβ/tau– vs. Aβ/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 assessments[63], showing their value as a screening tool for inclusion into 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 five years[32], while learning rates captured with C3 FNAME and

BRANCH showed associations with PACC change over one year[30,61]. 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 diseases[85].

Practical use of current tools and development of future tools

We found a total 23 tools that are used for remote and unsupervised data collection, 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. 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+ Framework[44]).

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 the The Library of Digital Measurement Products[86]. All the tools in Table 1 except OCTAL and ORCA-LLT can be found in the library[e], and new digital tools are added regularly.

Conclusion

This scoping review identified 27 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 colleagues[7], 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), and individualized prognostics and risk assessment—especially as it pertains to those affected by AD.

[e] https://datacc.dimesociety.org/digital-measurement-library/

References

1. 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).

2. 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).

3. Food and Drug Administration. Early Alzheimer's Disease: Developing Drugs for Treatment. Guidance for Industry. (2024).

4. Hampel, H. *et al.* Designing the next-generation clinical care pathway for Alzheimer's disease. *Nat. Aging* **2**, 692–703 (2022).

5. Dubois, B. *et al.* Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* **9**, 1118–1127 (2010).

6. 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).

7. 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**, (2021).

8. 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).

9. Rentz, D. M. *et al.* Association of Digital Clock Drawing With PET Amyloid and Tau Pathology in Normal Older Adults. *Neurology* **96**, (2021).

10. Falleti, 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).

11. Fray, P. J., Robbins, T. W. & Sahakian, B. J. Neuorpsychiatyric applications of CANTAB. *Int. J. Geriatr. Psychiatry* **11**, 329–336 (1996).

12. 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).

13. 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).

14. 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).

15. Muurling, M. *et al.* Remote monitoring technologies in Alzheimer's disease: design of the RADAR-AD study. *Alzheimers Res. Ther.* **13**, 89 (2021).

16. 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).

17. 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).

18. 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).

19. Samaroo, A. *et al.* Diminished Learning Over Repeated Exposures (LORE) in preclinical Alzheimer's disease. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **12**, (2020).

20. 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).

21. Weizenbaum, E. L. *et al.* Using a digital assessment of multi-day learning curves to detect preclinical AD. *Alzheimers Dement.* **18**, e062194 (2022).

22. 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).

23. 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).

24. 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).

25. Sliwinski, M. J. Measurement-Burst Designs for Social Health Research. *Soc. Personal. Psychol. Compass* **2**, 245–261 (2008).

26. Stawski, R. S., MacDonald, S. W. S. & Sliwinski, M. J. Measurement Burst Design. in *The Encyclopedia of Adulthood and Aging* (ed. Whitbourne, S. K.) 1–5 (Wiley, 2015). doi:10.1002/9781118521373.wbeaa313.

27. 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).

28. 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).

29. 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).

30. 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).

31. Lim, Y. Y. *et al.* Association of deficits in short-term learning and Aβ and hippocampal volume in cognitively normal adults. *Neurology* **95**, e2577–e2585 (2020).

32. Berron, D. *et al.* Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease. *Alzheimers Dement.* alz.13919 (2024) doi:10.1002/alz.13919.

33. Lancaster, C. *et al.* Gallery Game: Smartphone-based assessment of long-term memory in adults at risk of Alzheimer's disease. *J. Clin. Exp. Neuropsychol.* **42**, 329–343 (2020).

34. 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).

35. 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).

36. 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).

37. Shehab, A. & Abdulle, A. Cognitive and autonomic dysfunction measures in normal controls, white coat and borderline hypertension. *BMC Cardiovasc. Disord.* **11**, 3 (2011).

38. 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).

39. 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).

40. International Telecommunication Union. Measuring digital development: Facts and Figures 2023. (2023).

41. Siddiqa, A., Karim, A. & Gani, A. Big data storage technologies: a survey. *Front. Inf. Technol. Electron. Eng.* **18**, 1040–1070 (2017).

42. 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).

43. Romansky, R. P. & Noninska, I. S. Challenges of the digital age for privacy and personal data protection. *Math. Biosci. Eng.* **17**, 5288–5303 (2020).

44. 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).

45. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan—a web and mobile app for systematic reviews. *Syst. Rev.* **5**, 210 (2016).

46. 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).

47. 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).

48. Weiner, M. W. *et al.* Brain health registry updates: An online longitudinal neuroscience platform. *Alzheimers Dement.* alz.13077 (2023) doi:10.1002/alz.13077.

49. 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.* (2024) doi:10.14283/jpad.2024.32.

50. Langford, O. *et al.* PREDICTING AMYLOID BURDEN TO ACCELERATE RECRUITMENT OF SECONDARY PREVENTION CLINICAL TRIALS. *J. Prev. Alzheimers Dis.* 1–6 (2020) doi:10.14283/jpad.2020.44.

51. 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.* 1–7 (2021) doi:10.14283/jpad.2021.36.

52. Leuzy, A. *et al.* REAL AD—Validation of a realistic screening approach for early Alzheimer's disease. *Alzheimers Dement.* 1–11 (2024).

53. 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).

54. 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.* (2023) doi:10.14283/jpad.2023.11.

55. 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).

56. Boots, E. A. *et al.* Continuous Associations between Remote Self-Administered Cognitive Measures and Imaging Biomarkers of Alzheimer's Disease. *J. Prev. Alzheimers Dis.* (2024) doi:10.14283/jpad.2024.99.

57. 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).

58. 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).

59. 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).

60. Shiffman, S., Stone, A. A. & Hufford, M. R. Ecological Momentary Assessment. *Annu. Rev. Clin. Psychol.* **4**, 1–32 (2008).

61. 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).

62. 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).

63. 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).

64. 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).

65. 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).

66. 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).

67. Muurling, M. *et al.* Feasibility and usability of remote monitoring in Alzheimer's disease. *Digit. Health* **10**, 20552076241238132 (2024).

68. Miller, M. J. *et al.* The ADNI4 Digital Study: A novel approach to recruitment, screening, and assessment of participants for AD clinical research. *Alzheimers Dement.* alz.14234 (2024) doi:10.1002/alz.14234.

69. 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).

70. ter Huurne, D. *et al.* User Experience of a (Semi-) Automated Cognitive Phone-Based Assessment Within a Memory Clinic Population. *Arch. Clin. Neuropsychol.* acae063 (2024) doi:10.1093/arclin/acae063.

71. Sliwinski, M. J. *et al.* Reliability and Validity of Ambulatory Cognitive Assessments. *Assessment* **25**, 14–30 (2016).

72. Thompson, L. I. *et al.* Psychometric and adherence considerations for high-frequency, smartphone-based cognitive screening protocols in older adults. *J. Int. Neuropsychol. Soc.* 1–9 (2024) doi:10.1017/S1355617724000328.

73. Rentz, D. M. *et al.* THE FEASIBILITY OF AT-HOME IPAD COGNITIVE TESTING FOR USE IN CLINICAL TRIALS. *J. Prev. Alzheimers Dis.* 1–5 (2015) doi:10.14283/jpad.2015.78.

74. 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).

75. 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).

76. 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).

77. 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).

78. Donohue, M. C. *et al.* The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-Related Decline. *JAMA Neurol.* **71**, 961 (2014).

79. 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).

80. 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).

81. 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).

82. Bilder, R. M. & Reise, S. P. Neuropsychological tests of the future: How do we get there from here? *Clin. Neuropsychol.* **33**, 220–245 (2019).

83. Cid, R. E. C. & Loewenstein, D. A. Salient Cognitive Paradigms to Assess Preclinical Alzheimer's Disease. *Neurotherapeutics* **19**, 89–98 (2022).

84. Singh, S. & Germine, L. Technology meets tradition: a hybrid model for implementing digital tools in neuropsychology. *Int. Rev. Psychiatry* **33**, 382–393 (2021).

85. 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).

86. 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.* (2024) doi:10.14283/jpad.2024.103.