Neuroscience and Biobehavioral Reviews 187 (2026) 106756

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Digital episodic memory tools for the identification of amnestic mild cognitive impairment: A systematic review and an exploratory meta-analysis

Ana Rita Moura [a,b,c,d,*], Daniela Magalhães [e], Carolina dos Santos [f], Filipe Marinho [g], Rui Henriques [h], G. Elliott Wimmer [a,i], Raymond J. Dolan [a,i], Isabel Fragata [c,j]

[a] Max Planck Centre for Computational Psychiatry and Ageing, University College London, London, United Kingdom
[b] CHRC, NOVA Medical School, Faculdade de Ciências Médicas, NMS, FCM, Universidade NOVA de Lisboa, Lisboa, Portugal
[c] iNOVA4Health, NOVA Medical School, Faculdade de Ciências Médicas, NMS, FCM, Universidade NOVA de Lisboa, Lisboa, Portugal
[d] Department of Psychiatry and Mental Health, Unidade Local de Saúde de Lisboa Ocidental, Lisbon, Portugal
[e] Department of Psychiatry, Hospital de Cascais, Cascais, Portugal
[f] Department of Mental Health, Hospital Júlio de Matos, Unidade Local de Saúde S. José, Lisbon, Portugal
[g] Unidade Local de Saúde Amadora-Sintra, Amadora, Portugal
[h] INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Portugal
[i] Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom
[j] Department of Neuroradiology, Unidade Local de Saúde S. José, NOVA Medical School, Lisbon, Portugal

ARTICLE INFO

Keywords:
Digital memory assessment
Episodic memory
AMCI
Cognitive screening
Alzheimer's disease
Early detection
Systematic review
Meta-analysis

ABSTRACT

Within our ageing population, the prevalence of Amnestic Mild Cognitive Impairment (aMCI) is rising, underscoring a need for early detection strategies to mitigate potential progression to dementia. Traditional pen-and-paper cognitive tests remain standard but often fail to capture subtle early impairments. Digital episodic memory tasks have emerged as a promising alternative, offering automated scoring, increased accessibility, and potentially better sensitivity to cognitive decline.

We systematically evaluated the diagnostic accuracy of digital episodic memory assessments for aMCI. Thirteen studies met inclusion criteria, with six (n = 510 participants) providing sufficient data for meta-analysis. Pooled estimates showed sensitivity of 0.78 (95% CI: 0.65–0.87), specificity of 0.79 (95% CI: 0.66–0.89), and AUC of 0.86, with low to moderate heterogeneity ($I^2 = 36.1\%$), largely reflecting differences in task design and administration.

Both quantitative and qualitative findings suggest digital episodic memory tasks can provide accessible, scalable, and effective tools for early aMCI detection. However, design variability and construct differences highlight the need for further validation across diverse populations and improved methodological standardisation before clinical implementation.

1. Introduction

Amnestic Mild Cognitive Impairment (aMCI) represents an intermediate stage between normal ageing and dementia, characterised by episodic memory deficits linked to a significant increase in risk of progression to Alzheimer's disease (AD) (Albert et al., 2013; Petersen, 2004). The prevalence of aMCI increases with age, affecting approximately 6.7% of individuals aged 60–64 and rising to 25.2% in those over 80 (Lawrence et al., 2017). While not all individuals with aMCI develop AD, those with positive biomarkers, such as amyloid-β (Aβ) aggregation, show an accelerated risk of conversion (Alves et al., 2021; Wolk et al., 2009).Given the increasing availability of disease-modifying treatments targeting early AD pathology, an important clinical priority is to identify individuals with aMCI who are most likely to progress to AD. Reliable discrimination between aMCI and healthy ageing nevertheless remains a necessary upstream step in this process. In clinical settings, memory complaints are common in older adults and often reflect non-pathological ageing or affective conditions rather than

* Correspondence to: Campo dos Mártires da Pátria 130, Lisboa 1169-056, Portugal.
E-mail address: rita.moura@nms.unl.pt (A.R. Moura).

https://doi.org/10.1016/j.neubiorev.2026.106756

Available online 25 May 2026

neurodegenerative disease (Petersen, 2016). Accurate cognitive discrimination therefore serves as an initial triage step to identify individuals who may warrant further biomarker assessment, longitudinal monitoring, or inclusion in prognostic studies (Hansson, 2021).

Currently, aMCI diagnosis relies on a combination of clinical interviews, neuropsychological testing, and potentially biomarker evaluation (Petersen, 2016). Of all cognitive tests, performance on episodic memory tasks consistently demonstrates predictive value for progression from aMCI to AD, underscoring a need for refined, behaviour-based, diagnostic tools (Gustavson et al., 2020; Summers and Saunders, 2012). Additionally, treatment is recognized as effective when it is associated with changes in clinically meaningful endpoints, whether cognitive or functional, underscoring a need for measures that can reliably capture such changes.

Traditional pen-and-paper neuropsychological assessments suffer many limitations, including low sensitivity to early-stage impairments, cultural and educational biases, lengthy administration times, and susceptibility to practice and ceiling effects (Kourtis et al., 2019). While biomarker-based approaches (e.g., cerebrospinal fluid analysis and amyloid PET imaging) offer valuable diagnostic insights, their high cost, invasiveness, and limited accessibility restrict their widespread use in routine clinical settings (Ossenkoppele et al., 2022). These constraints highlight an urgent need for scalable, accessible, and non-invasive diagnostic tools to detect aMCI before significant cognitive decline ensues.

Digital/computerised cognitive assessments have emerged as one potential solution, leveraging computerised episodic memory tasks to enhance sensitivity, accessibility, and automation in detecting subtle cognitive impairments. Several commercially available platforms, including CogState (Cogstate: Simplifying the Measurement of Cognition, n.d.), CAMCI (Saxton et al., 2009), CANTAB (Égerházi et al., 2007), CNS Vital Signs (Gershon et al., 2010), and the NIH Cognition Battery (Gershon et al., 2010), in effect digitise traditional neuropsychological tasks, while offering real-time scoring and adaptive difficulty adjustments. However, evidence regarding their diagnostic accuracy is mixed, and their ability to discriminate early-stage aMCI from healthy ageing has not been systematically assessed.

Beyond traditional digital testing, newer findings in cognitive neuroscience have sought to refine episodic memory assessment by focusing on relational memory, temporal context, and multi-step decision-making, features that may offer a better proxy of hippocampal dysfunction in early aMCI (Delhaye et al., 2019a, 2019b; Guillery-Girard et al., 2013; Shadlen and Shohamy, 2016; Shohamy and Daw, 2015). Tasks that assess associative and relational memory, rather than simple recall may offer improved sensitivity by capturing deficits that precede standard episodic memory impairments(Gershon et al., 2010). However, these advanced paradigms, and their translational potential, have yet to be validated fully in clinical populations.

No prior systematic review has evaluated the diagnostic efficacy of digital episodic memory assessments for aMCI, nor synthesised their differences in sensitivity and specificity against reference traditional methods.

The literature search focuses on studies of episodic memory, defined broadly to include episodic retrieval, associative memory, event memory, and context memory. Studies using spatial memory tasks are included when the task is conceptually linked to episodic processes (e.g., spatial context learning, navigation within episodic paradigms, or maze tasks testing event-place associations), recognizing the close neurocognitive relationship between episodic and spatial memory (Shohamy and Daw, 2015). This review characterises such tasks and evaluates their ability to distinguish cognitively unimpaired (CU) older adults from individuals with aMCI. In doing so, we organise the literature along a conceptual continuum ranging from direct digital translations of traditional episodic memory tests to newer paradigms grounded in contemporary cognitive neuroscience models. By synthesising available evidence, this study aims to determine whether digital episodic memory tasks offer a reliable and scalable complement to traditional neuropsychological assessments and to identify gaps in current methodologies that warrant further research.

2. Methods

2.1. Search strategy

This review was registered with PROSPERO (CRD42024622135) and conducted in accordance with the Preferred Reporting Items for Systematic Review and Meta-analyses (PRISMA-DTA) guidelines (McInnes et al., 2018; Salameh et al., 2020) (Supplementary Table 1). Two authors (ARM, CS) searched the following databases PubMed, Scopus, Web of Science, Cochrane Library, IEEE, ACM for retrospective and prospective cohort studies, cross-sectional studies, and randomised clinical trials that use episodic memory digital tasks to diagnose patients with cognitive decline, cognitive impairment from each database's inception to 11 of April 2025. The full search strategy is included in Supplementary Tables 3–7 in the Supplement. Two authors (ARM, CS) independently screened abstracts and titles followed by full texts to check the eligibility for inclusion, with disputes being resolved through consensus from a third independent author (IF) OR discussion. Article screening was done using the online platform Rayyan (Ouzzani et al., 2016).

The methodology and reporting of this review followed guidance set out by the Cochrane Handbook for Diagnostic Test Accuracy Reviews. The keywords advanced search used were ("computational task*" or "Computer-assisted" or "computational" or "digital" or "web-based" or "mobile application" or "smartphone application" or "mobile" or "smartphone" or "app" or "apps" or "application*" or "computer*" or "computational test*" or "digital biomarker") AND (diagnos* or Stratifi* or Classifi* or Assess* or predict* or detect* or identif*) AND ("Amnestic Mild Cognitive Impairment" or "amnestic MCI" or "aMCI" or "Aging" or "ageing" or "older adults" or "elderly" or "amnestic Cognitive Dysfunction" or "amnestic memory cognitive impairment" or "amnestic cognitive impairment" or "amnestic cognitive decline") AND ("Episodic Memory" or "episodic retrieval" or "associative memory" or "event memory" or "autobiographical memory" or "context memory" or "recollection" or "memory for events" or "memory for experiences").

2.2. Eligibility

The inclusion criteria of this study are as follows: 1) Original research article, 2) An observational study conducting a qualitative and/or quantitative evaluation of the diagnostic performance in a digital/computerised episodic memory test for the identification of aMCI vs healthy ageing, 3) Healthy participants and participants with aMCI diagnosed with standardised diagnostic criteria, including the Petersen criterion, the report of the International Working Group on Mild Cognitive Impairment (Winblad et al., 2004), the recommendations of the National Institute on Aging and the Alzheimer's Association, the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS), the Alzheimer's Disease and Related Disorders Association, any version of the Diagnostic and Statistical Manual of Mental Disorder, consensus by qualified clinicians, or standardised neuropsychological tests like the MoCA 4) Healthy/control participants assessed as cognitively healthy by clinicians, CDR, or standardised neuropsychological tests, 5) To be included in the quantitative meta-analysis the study should further report or allow for an estimate of diagnostic performance in terms of sensitivity and specificity, or provide sufficient data to derive these metrics. Studies lacking the necessary performance information were considered for qualitative review but excluded from the quantitative meta-analysis.

The exclusion criteria were as follows: 1) Languages other than English, Spanish, French, Italian, and Portuguese, 2) The digital test was administered for diseases other than aMCI or cognitively healthy

participants, such as patients with brain injury, 3) Patients with other psychiatric comorbidities, 4) The digital test was not a cognitive test, 5) The digital test's primary cognitive target was a domain other than episodic memory (e.g., prospective memory or metamemory paradigms), even when episodic processes were known to contribute. 5) Tests delivered via immersive platforms were excluded. Desktop- or laptop-class computers, including laboratory touchscreen systems, were eligible. Smartphone-implementations were excluded a priori and 6) Case studies, commentaries, editorials, and reviews.

The selection of studies for this review included those that utilised computerised tasks designed to assess episodic memory and provided diagnostic performance data, such as sensitivity and specificity. Attention was given to a subset of studies that incorporated Receiver Operating Characteristic (ROC) curve analysis and an associated AUC value (as a proxy statistic for both effect size and well-calibrated decision's confidence) that evaluated an ability to distinguish accurately between amnestic Mild Cognitive Impairment (aMCI) and healthy adults. Studies reporting diagnostic accuracy metrics (e.g., sensitivity, specificity, AUC) were included in the quantitative meta-analysis, whereas experimental paradigms without such metrics were included in the qualitative synthesis only.

2.3. Data extraction and study characteristics

Relevant data were extracted into a standardised template independently by two reviewers (ARM, CS), covering author, year, study design, sample size, index test, comparator tests, setting, participant demographics, diagnostic criteria, and technical details of the index tests (e.g., platform, administration method, duration, tester expertise). Diagnostic efficacy metrics for meta-analysis were also extracted. Summary tables (Tables 1–3) provide an overview of study populations, interventions, test platforms, and diagnostic outcomes. A third reviewer (DM) verified data extraction. For studies lacking confusion matrix data these were reconstructed from reported sensitivity, specificity, and case-control sample sizes.

2.4. Risk of bias assessment

Quality assessment was conducted independently by two investigators (ARM, CS) and disagreements were resolved by consensus or appeal to a third author (FM). with the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) checklist (Whiting et al., 2011). QUADAS-2 evaluates risk of bias across four core domains: Patient selection assesses whether study participants were recruited in a way that could introduce bias, including the use of case–control designs, inappropriate exclusion criteria, or non-representative sampling. Index test evaluates whether the digital memory test under investigation was conducted and interpreted without knowledge of the reference standard results, and whether diagnostic thresholds were pre-specified rather than determined post hoc. Reference standard examines whether the method used to classify aMCI (e.g., neuropsychological assessment, clinical diagnosis) is likely to correctly identify the target condition, and whether its interpretation was blinded to the index test. Flow and timing evaluate whether all participants received both the index test and the reference standard, whether exclusions were appropriate, and whether the time interval between tests was sufficiently short to minimise disease progression bias. Each domain was rated as having low, high, or unclear risk of bias in accordance with QUADAS-2 guidance. Applicability concerns were assessed for the first three domains (patient selection, index test, and reference standard) to evaluate how well included studies matched the review question with respect to population, test characteristics, and diagnostic criteria.

2.5. Data synthesis and meta-analysis

The statistical analysis followed the Cochrane Guidelines for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0 (Bossuyt et al., 2023) and the Methods Guide for Authors of Systematic Reviews of Medical Tests by the Agency for Healthcare Research and Quality (chapter 8) (Lipsey, 2001).

2.5.1. Diagnostic accuracy meta-analysis

We summarized study characteristics in structured tables (Tables 1–3) and performed meta-analyses when at least three studies reported comparable diagnostic metrics (e.g., sensitivity, specificity, AUC). For each study, we computed the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) counting using data from the index test (Doebler and Holling, 2015). For studies that did not explicitly report confusion matrix data, the true positives, false positives, true negatives, false negatives were recovered from reported sensitivity, specificity, sample sizes, and prevalence when available (Doebler and Holling, 2015).

For studies reporting only model-based estimates from composite scores combining multiple cognitive domains (e.g., memory, executive, attention), we qualitatively summarized findings but excluded them from the meta-analysis. This approach ensured comparability across studies focusing specifically on digital episodic memory tests.

We used random-effects models with 95% confidence intervals and assessed heterogeneity using I^2, interpreting 25%, 50%, and 75% as low, moderate, and high heterogeneity (Higgins et al., 2003). Publication bias was investigated using Deeks' funnel plot by conducting a sample size-weighted regression of the log odds ratio against the inverse of the square root of the sample size (Deeks et al., 2005). Analyses were conducted using mada(Schwarzer et al., 2015) and metafor package (version 4.6.0) (Viechtbauer, 2010) in R. Statistical tests are two-sided, and the confidence threshold for statistical significance at $p < 0.05$.

2.6. Certainty of evidence (GRADE)

Certainty of evidence for pooled outcomes was evaluated using GRADE criteria, considering risk of bias, inconsistency, indirectness, imprecision, and publication bias. Overall certainty was rated as moderate.

2.7. Synthesis of results

When synthesising results, studies that reported different thresholds for defining positive test results are highlighted. When studies included indeterminate or missing test results, we followed what the authors of the study had defined. For the inclusion of spatial memory tasks, we focused only on tasks that were conceptually linked to episodic processes, such as spatial context learning or maze navigation with event-place associations. We explicitly excluded purely working memory or spatial memory tasks without an episodic memory component to stay aligned with our primary research question.

3. Results

3.1. Study selection

A systematic review following PRISMA-DTA 2020 guidelines and the established inclusion criteria was conducted. Studies were eligible if they evaluated digital or computerised memory tests in older adults with aMCI compared to HC or other cognitive groups. We included studies reporting test characteristics, sample demographics, cognitive domains, and diagnostic outcomes. Sample sizes and diagnostic classification counts (TP, FP, TN, FN) per study are presented in supplementary Table 5.

A systematic search yielded 208 potentially relevant articles. After deleting the deduplicates, 135 articles were retained for further review, with three additional reports identified through citation tracking (Junkkila et al., 2012; Paterson et al., 2022; Wang et al., 2013). Title and

Table 1
Characteristics of Included Studies. Summary of included studies evaluating digital cognitive assessments for amnestic Mild Cognitive Impairment (aMCI). This table presents key characteristics of 13 studies included in the systematic review. It outlines the digital cognitive test used, sample composition, geographic and clinical setting, mean age of the aMCI group, targeted cognitive domain, and whether diagnostic accuracy metrics (e.g., sensitivity, specificity, area under the curve [AUC]) were reported.

Author, Year	Country/ Setting	Study Design	Sample Size (Groups)	Mean Age (SD, Range)	Gender (% Male)	Mean Education (SD, Range)	Condition (s)	aMCI Criteria	Digital Index Test	Cognitive Domain	Reference/ Traditional Tests	Diagnostic Metrics Reported	Method of Administration	Details of Test Platform Used	Time (min)
Alegret et al. (2020)	Spain, Memory clinic (Fundació ACE, Barcelona)	Cross-sectional	276 total (HC = 154, aMCI = 61, naMCI = 61)	aMCI: 67.7 (7.9)	aMCI: 37.7% male (23/61)	aMCI: 10.2 (4.0) years	HC, naMCI, aMCI	Petersen criteria; CDR = 0.5; MMSE ≥ 24; preserved ADLs; objective memory impairment	FACEmemory® (computerized Face–Name Associative Memory Exam, short form)	Associative episodic memory (face–name–occupation)	MMSE; Neuropsychological Battery of Fundació ACE (NBACE); S-FNAME (for convergent validity)	Sensitivity, specificity, AUC	Self-administered on tablet with minimal supervision	Tablet-based application with touchscreen and voice recognition; automatic scoring; Android-based custom software	~30 min
Belliart-Guérin and Planche (2023)	France, Hospital clinic	Retrospective	90 (SCC, aMCI, naMCI, dementia)	aMCI:67 (49–76)	aMCI:53%	aMCI:12 (5–22)	SCC, naMCI, aMCI, dementia	DSM-5, Petersen	Mnemonic Similarity Task	Episodic memory	MMSE, FAB	Sensitivity, specificity, AUC	Individual administration, minimal supervision	Computer, details not specified	13
Junkkila et al. (2012)	Finland, Neurology clinic	Cross-sectional	58 (HC, aMCI, AD)	aMCI:73 (6.3)	aMCI:65	aMCI:8 (3.0)	HC, aMCI, AD	Petersen et al., 2001	CANTAB PAL	Visual-spatial associative	WLMT	AUC	Individual, supervised administration	Touchscreen computer	Not specified
Kroft et al. (2022)	Canada, Hospital	Cross-sectional	40 (HC, aMCI)	aMCI:65.8 (11.1)	Not reported	aMCI:16.2 (3.2)	HC, aMCI	DSM-5	OST Memory Task	Episodic memory	None	No	Individual administration	Computer + Serial Response Box (E-Prime software)	Not specified
Lee et al. (2014)	South Korea, Hospital	Cross-sectional, longitudinal	60 (HC, aMCI, AD)	aMCI:70.7 (5.0)	aMCI:45%	aMCI:11.0 (4.2)	HC, aMCI, AD	Petersen et al., 1999	Virtual Radial Arm Maze	Spatial working/ reference	None (neuropsych tests used only for correlation)	No	Individual administration	Computer-based, joystick-operated virtual environment	Not specified
Paterson et al. (2022)	Canada, Community	Cross-sectional	91 (HC, aMCI)	aMCI:75 (5.7)	aMCI:51%	aMCI:15 (2.5)	HC, aMCI	NIA-AA criteria	Brain Health Assessment	Episodic/ working memory	MoCA	Sensitivity, specificity, AUC	Individual, unsupervised administration	Laptop with mouse	20–30
Pirogovsky et al. (2013)	USA, Research center	Cross-sectional	30 (HC, aMCI)	aMCI:76.8 (2.3)	Not reported	aMCI:15 (0.9)	HC, aMCI	Petersen et al., 2001	Temporal Sequence Learning Task	Temporal sequence memory	None	No (group differences only)	Individual administration	Not specified	Not specified
Ramratan et al. (2012)	USA, Community-based	Cross-sectional	390 (HC, aMCI)	aMCI:81.7 (5.5)	aMCI:53%	aMCI:13.1 (3.4)	HC, aMCI	Petersen et al., 1999	CRRST	Cued-recall speed	FCSRT, WMS-R	AUC (no cut-off, no binary stats)	Individualadministration with research assistant present	Computer + E-Prime program; oral responses with mic	Not specified
Rigby et al. (2024)	USA, Community-dwelling older adults (Michigan Alzheimer's Disease Research Center)	Cross-sectional	158 total (HC = 96; aMCI = 62)	aMCI(72.1 (6.8))	aMCI: 24 (38.7%)	aMCI: 15.6 (2.4)	HC and aMCI	NACC consensus diagnosis (Unified Data Set v3); amnestic features required	NIH Toolbox Cognition Battery (NIHTB-CB), tablet version Primary discriminant subtest: Picture Sequence Memory	Episodic memory (primary)	NACC Unified Data Set neuropsychological battery (used for consensus diagnosis, not as comparator test)	No	Individual administration on tablet (iPad), supervised research setting	NIHTB-CB tablet version; norm-adjusted and unadjusted scores analysed	~30 min

(*continued on next page*)

Table 1 (*continued*)

Author, Year	Country/ Setting	Study Design	Sample Size (Groups)	Mean Age (SD, Range)	Gender (% Male)	Mean Education (SD, Range)	Condition (s)	aMCI Criteria	Digital Index Test	Cognitive Domain	Reference/ Traditional Tests	Diagnostic Metrics Reported	Method of Administration	Details of Test Platform Used	Time (min)
Servais, Barbeau and Bastin (2024)	Belgium/ France, Clinic	Cross-sectional	43 (HC, aMCI)	aMCI:69.9 (7.1)	aMCI:43%	aMCI:13.7 (3.3)	HC, aMCI	Petersen and Negash, 2008	Oddball & Von Restorff Paradigms	Novelty detection memory	None (classification based on logistic model)	Sensitivity, specificity	Visual tasks, manual keypress, oral recall	Computer (OpenSesame software)	Not specified
Troyer et al. (2016)	Canada, Geriatric unit	Cross-sectional	45 (HC, aMCI)	aMCI:76.1 (7.6)	aMCI:37%	aMCI:14.2 (2.5)	HC, aMCI	Knopman et al., 2003	Associative Recognition	Associative memory	None	No	Individual, self-paced with keypress responses	Computer, custom software	Not specified
Wams et al. (2017)	UK, Hospital	Cross-sectional	45 (HC, aMCI, AD)	aMCI:77.1 (4.0)	aMCI:50%	Not reported	HC, aMCI, AD	Petersen et al., 1999	CANTAB Eclipse Battery	Visual/ verbal memory	None	No	Individual administration in home setting	Touchscreen computer (CANTAB Eclipse)	Not specified
Wang et al. (2013)	Taiwan, Hospital	Cross-sectional	90 (HC, aMCI, mild dementia)	aMCI:73.7 (7.4)	aMCI:47%	aMCI:8.4 (3.4)	HC, aMCI, m-DAT	Petersen et al., 1999 + battery	Spatial-Context Memory Test	Spatial, associative memory	WLMT	Sensitivity, specificity, AUC	Individual administration	Computer	15–20

AUC: Area Under the Curve, SCC: Subjective Cognitive Complaint, naMCI: Non-Amnestic Mild Cognitive Impairment, aMCI: Amnestic Mild Cognitive Impairment, CH: Cognitively Healthy, MCI: Mild Cognitive Impairment, MMSE: Mini-Mental State Examination, FAB: Frontal Assessment Battery, PAL: Paired Associate Learning, CERAD: Consortium to Establish a Registry for Alzheimer's Disease, CRRST: Cued-Recall Retrieval Speed Task, FACEmemory®: Face-Name Associative Memory Test.

abstract screening reduced the pool to 17 articles for full-text review, ultimately resulting in thirteen studies (Belliart-Guérin and Planche, 2023; Alegret et al., 2020, Rigby et al., 2024; Junkkila et al., 2012; Kroft et al., 2022; Lee et al., 2014;Paterson et al., 2022; Pirogovsky et al., 2013; Ramratan et al., 2012; Servais et al., 2024; Troyer et al., 2016; Wams et al., 2017 and Wang et al., 2013) included in the qualitative synthesis. Of these, six studies were eligible for the meta-analysis (Alegret et al., 2020; Belliart-Guérin and Planche, 2023; Junkkila et al., 2012; Paterson et al., 2022; Servais et al., 2024; Wang et al., 2013); the remaining studies were excluded due to missing or insufficient diagnostic data. Three studies were excluded at the full-text screening stage: The study by (Pagán et al., 2023) was excluded due to its focus on attention and executive function via a go/no-go error--monitoring paradigm rather than episodic memory; Chi et al. (2014) was excluded due to its primary focus on prospective memory, a cognitive construct conceptually distinct from episodic memory; Planche, 2023 was excluded because it used a paper-based booklet with computer-generated images; LaPlume et al. (2021) was excluded due to sample overlap with Paterson et al., 2022, identified during citation tracking; the latter was retained as the primary reference. The PRISMA flow diagram (Fig. 1) provides a detailed overview of the study selection process.

3.2. Study characteristics

Thirteen studies evaluating digital tools for assessing episodic memory in aMCI populations were included (Tables 1–2). Most used cross-sectional designs, with sample sizes ranging from 30 to 390 participants. Participant groups included healthy controls or individuals with subjective memory complaints (Belliart-Guérin and Planche, 2023) and individuals with aMCI, with mean ages ranging from 65.8 to 81.7 years. aMCI diagnosis was based on established criteria, including Petersen (Petersen, 2004), National Institute on Aging – Alzheimer's Association (NIA-AA)(Albert et al., 2011) or DSM-5 definitions (Diagnostic and statistical manual of mental disorders, 2013).

The studies encompassed a range of settings, including memory clinics, neurology departments, and community samples, across Europe, North America, and Asia. A variety of episodic memory domains were assessed, including episodic, associative, and spatial memory as can be consulted in Tables 1–3. Spatial memory tasks were analysed alongside episodic memory tasks when they measured overlapping cognitive processes relevant to aMCI.

3.3. Results of individual studies

Table 3 summarizes index test characteristics, platforms, administration methods, durations, and tester expertise. Tests were implemented on computers, touchscreens, or joystick-controlled computers. Administration methods ranged from fully self-administered tasks with minimal supervision (e.g., Brain Health Assessment) to supervised delivery by neuropsychologists or trained research staff. Test durations varied from ~10–45 min, while administration modes varying from self-administered tasks to fully supervised assessments by trained personnel.

Different episodic memory tasks were used where Pirogovsky et al. (2013) assessed temporal sequence learning through a visuospatial task in which participants had to memorise and reproduce fixed sequences of circles on a maze. Wams et al. (2017) used selected tasks from the CANTAB battery to evaluate visual and episodic memory, visuospatial memory, and verbal object recognition. Wang et al. (2013) administered a modified spatial context memory test comprising three subtests: spatial location, event-place association, and place-object association, to evaluate spatial memory based on real-world mapping and object-scene pairing. Troyer et al. (2016) employed two associative recognition tests, one using word–word pairs and another using face–name pairs, to investigate recognition memory with accuracy and reaction time measures. Servais et al. (2024) examined novelty detection

Table 2
Summary of Diagnostic Accuracy Metrics for Digital Episodic Memory Tests in Identifying aMCI. AUC: Area Under the Curve, SCC: Subjective Cognitive Complaint, naMCI: Non-Amnestic Mild Cognitive Impairment, aMCI: Amnestic Mild Cognitive Impairment, CH: Cognitively Healthy, MCI: Mild Cognitive Impairment, MMSE: Mini-Mental State Examination, FAB: Frontal Assessment Battery, PAL: Paired Associate Learning, CERAD: Consortium to Establish a Registry for Alzheimer's Disease, CRRST: Cued-Recall Retrieval Speed Task, FACEmemory®: Face-Name Associative Memory Test. NA = Not available; data were not reported or could not be derived from the publication.

Author, Year	Index Test	Cut-off / Classification Metric	Sensitivity (%)	Specificity (%)	AUC
Belliart-Guérin and Planche (2023)	Mnemonic Similarity Task	Lure discrimination index 12	87	64	0.77
Junkkila et al. (2012)	CANTAB PAL	Total errors adjusted > 42	71	95	0.803
Junkkila et al. (2012)	CANTAB PAL and CERAD as covariate	Total errors adjusted > 42 and score > 6	76	95	0.906
Paterson et al., 2021	Brain Health Assessment	Single ROC-based cut-point	57	83	0.76
Servais, Barbeau and Bastin (2024)	Oddball Paradigms/ Target Stimuli RT	Logistic regression	71	59	0.61
Servais, Barbeau and Bastin (2024)	Von Restorff Paradigms/Gain Factor (90-pt target)	Logistic regression	33	86	0.56
Wang et al. (2013)	Spatial-Context Memory Test	Total score ≤ 15	97	93	0.99
Alegret et al. (2020)	FACEmemory®	Total score ≤ 31.5	80.5	80	0.85

Table 3
Confusion matrix outcomes for each study reviewed. TP = true positives; FN = false negatives; TN = true negatives; FP = false positives. AUC: Area Under the Curve, SCC: Subjective Cognitive Complaint, naMCI: Non-Amnestic Mild Cognitive Impairment, aMCI: Amnestic Mild Cognitive Impairment, CH: Cognitively Healthy, MCI: Mild Cognitive Impairment, MMSE: Mini-Mental State Examination, FAB: Frontal Assessment Battery, PAL: Paired Associate Learning, CERAD: Consortium to Establish a Registry for Alzheimer's Disease, CRRST: Cued-Recall Retrieval Speed Task, FACEmemory®: Face-Name Associative Memory Test. * Participants with AD and aMCI were grouped for the purpose of disease classification analyses.

Author, Year	TP	FN	TN	FP	aMCI (n)	HC (n)
Belliart-Guérin and Planche (2023)	26	4	21	12	30	33
Alegret et al. (2020)	49	12	123	31	61	154
Junkkila et al. (2012) (PAL only)	12*	5	21	1	17	22
Junkkila et al. (2012) (PAL and CERAD as covariate)	13*	4	21	1	17	22
Paterson et al. (2021)	29	22	33	7	51	40
Servais, Barbeau and Bastin (2024) (Oddball)	15	6	13	9	21	21
Servais, Barbeau and Bastin (2024) (Von Restorff Paradigms/Gain Factor (90-pt target))	7	17	18	3	21	21
Wang et al. (2013)	29	1	28	2	30	30

using an oddball paradigm and a Von Restorff task to evaluate attentional responses to novel versus repeated stimuli. Ramratan et al. (2012) used a verbal learning task with cued recall and reaction time measurement, requiring participants to match category cues with correct exemplars. Paterson et al. (2022) applied the online Brain Health Assessment (BHA), which includes tasks measuring spatial working memory, executive function, and associative recognition Alegret et al. (2020) introduced FACEmemory®, a tablet-based associative memory task focused on face–name learning and recall, designed as a digital screening tool for early Alzheimer's disease and amnestic MCI.

Lee et al. (2014) used a virtual radial-arm maze (VRAM) to assess working and reference memory through spatial navigation and reward retrieval. Kroft et al. (2022) recreated a memory task originally designed by Babb and Johnson (2010) to test object, spatial, and temporal memory through visual scene recall. Junkkila et al. (2012) employed the CERAD Word List Learning Task for verbal episodic memory and the CANTAB Paired Associate Learning (PAL) task to assess visual associative learning across increasing levels of difficulty. Finally, Belliart-Guérin and Planche (2023) applied the Mnemonic Similarity Task to evaluate pattern separation abilities by distinguishing between repeated, novel, and similar visual stimuli. Rigby et al. (2024) used the tablet-based NIH Toolbox Cognition Battery, with a primary focus on the Picture Sequence Memory Test, to assess episodic memory under standardised, norm-referenced conditions in community-dwelling older adults.

3.4. Risk of bias and applicability

Risk of bias was assessed using the QUADAS-2 tool, which showed generally moderate to high risk across studies. Most studies applied appropriate patient selection and valid reference standards, but several lacked blinding in index test interpretation or prespecified diagnostic thresholds. Although task administration was typically automated and unsupervised, risk of bias in the index test domain mainly reflected post-administration processes, including data cleaning, threshold derivation, feature selection, and analytic decisions that were not always clearly independent of diagnostic status. A detailed risk of bias is provided in Figs. 4 and 5.

3.5. Diagnostic accuracy and exploratory meta-analysis

Given the limited number of eligible studies and the heterogeneity in task design, populations, and reference standards, the following meta-analysis is presented as an exploratory pooled estimate intended to characterise the current state of evidence rather than to provide a definitive diagnostic benchmark. Pooled estimates should not be used to guide clinical decisions at this stage. Six studies provided sufficient diagnostic data for inclusion in the quantitative synthesis. Diagnostic efficacy was analysed using forest plots to visualise sensitivity and specificity across studies (Fig. 2). The six unique studies included a total of 510 participants, with sample sizes ranging from 39 to 215 per study. Across studies, both aMCI and healthy control participants were enrolled with varying group compositions (Table 1). Where diagnostic accuracy metrics were not reported, authors were contacted to request additional data. Additional estimates were obtained for one study (Servais et al., 2024), while no further information was received for the remaining studies during the revision period.

Sensitivity ranged from 0.57 (95% CI: 0.43–0.70) to 0.97 (95% CI: 0.80–1.00); specificity ranged from 0.59 (95% CI: 0.38–0.77) to 0.93 (95% CI: 0.77–0.98). The area under the summary receiver operating characteristic (SROC) curve was 0.84, indicating good overall discriminative ability (Figs. 3 and 4).

A bivariate random-effects meta-analysis (Reitsma method) yielded a pooled sensitivity of 0.78 (95% CI: 0.65–0.87) and specificity of 0.79 (95% CI: 0.66–0.89), accounting for between-study variability. Between-study heterogeneity was low to moderate, with an I^2 estimate of 36.1% using the Zhou and Dendukuri approach, and between 53.7–69.8% and 3.8–5.8% using Holling's unadjusted and adjusted approaches (Doebler and Holling, 2015).

The observed pooled sensitivity and specificity of episodic memory tasks in discriminating aMCI from other groups (e.g., healthy controls, mild dementia) were comparable to published estimates for widely used screening measures such as the MMSE and MoCA (Li et al., 2018; Malek-Ahmadi and Nikkhahmanesh, 2024), suggesting that episodic memory paradigms may offer a viable digital alternative for cognitive screening. The certainty of evidence was rated as moderate using the GRADE framework, primarily due to concerns regarding inconsistency

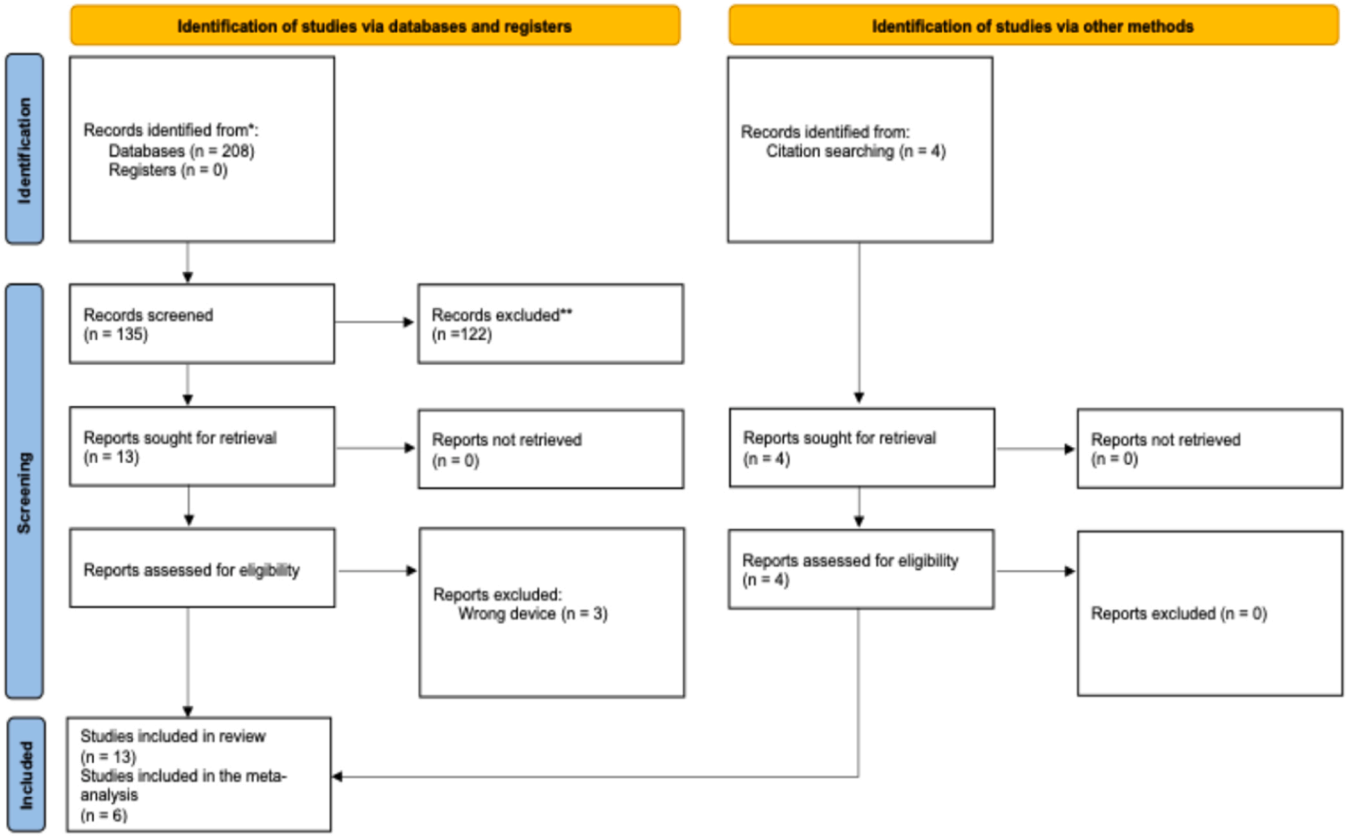

Fig. 1. PRISMA 2020 Flow Diagram of study selection and inclusion.

(moderate heterogeneity) and risks of bias (lack of blinding in some studies) (Supplementary Table S4).

4. Discussion

4.1. Cross-study considerations

This systematic review and meta-analysis serve as a guiding framework for understanding diagnostic capacity of current digital episodic memory assessment tools for detecting aMCI. Pooled sensitivity was 0.78 (95% CI: 0.65–0.87), specificity of 0.79 (95% CI: 0.66–0.89), and AUC of 0.86, providing an exploratory pooled estimate of diagnostic performance across a small and heterogeneous set of studies. Heterogeneity was low to moderate ($I^2 = 36.1\%$), although all studies were rated as having high risk of bias, which warrants cautious interpretation and precludes treating these pooled estimates as a stable diagnostic benchmark.

The literature includes both purpose-built diagnostic tools and experimental computerized paradigms originally developed to investigate memory mechanisms. While the latter were not designed as clinical diagnostic tools, they provide insight into cognitive processes relevant to early aMCI detection. This distinction is important when interpreting the evidence synthesised here: some digital tools aim primarily to operationalise clinical screening, whereas others explore mechanistically informed paradigms that may represent earlier stages in the development of diagnostic assessments. The heterogeneity observed across studies likely reflects the early developmental stage of digital cognitive assessment for aMCI, where rapid digital implementation has often preceded theoretical integration and clinical validation.

Despite wide variation in task design, administration, and scoring thresholds, one consistent finding emerged: tools grounded in hippocampal-dependent memory processes, particularly spatial context episodic and relational memory, tended to show stronger diagnostic utility than broader cognitive screeners. These processes are thought to be relevant for episodic memory function and might show earlier signs of impairment in Alzheimer's disease (Eichenbaum, 2017; Jack et al., 2000), aligning them with the goals of early detection and recent calls to re-centre behaviour in neuroscience to bridge mechanistic advances with clinically meaningful outcomes (Cirulli and Easton, 2022).

Consistent with this, meta-analytic evidence from non-digital cognitive assessments reports an overall AUC of 0.84 (95% CI: 0.81–0.87), alongside large group-level effects (Hedges' $g \approx 1.5$), indicating substantial separation between aMCI and cognitively unimpaired individuals at the group level (Malek-Ahmadi and Nikkhahmanesh, 2024). The pooled diagnostic accuracy observed here (SROC AUC $\approx$ 0.87) therefore falls within a similar range. This comparison is intended to provide contextualisation rather than direct equivalence, given differences in task structure, populations, and reference standards.

Across studies, two broad categories of digital assessments could be distinguished: (i) digital implementations of traditional neuropsychological paradigms and (ii) novel paradigms designed to probe specific episodic memory mechanisms such as associative binding or pattern separation. These approaches differ both in theoretical grounding and level of clinical validation. The included studies demonstrate that specialised digital paradigms hold promise:

- Belliart-Guérin et al. (2023) employed the Mnemonic Similarity Task (MST), a pattern separation paradigm where participants view objects during an initial encoding phase and later must classify each item in a three-alternative forced-choice format: "old," "lure," or "new". This test yielded moderate accuracy in distinguishing aMCI from subjective cognitive complaints (SCC) with a lure discrimination index cut-off of 12, sensitivity reached 87% and specificity 64% (AUC = 0.77), comparable to the MMSE (AUC = 0.77). This suggests that MST, despite being a brief digital tool, may match conventional screens for detecting subtle memory impairment.

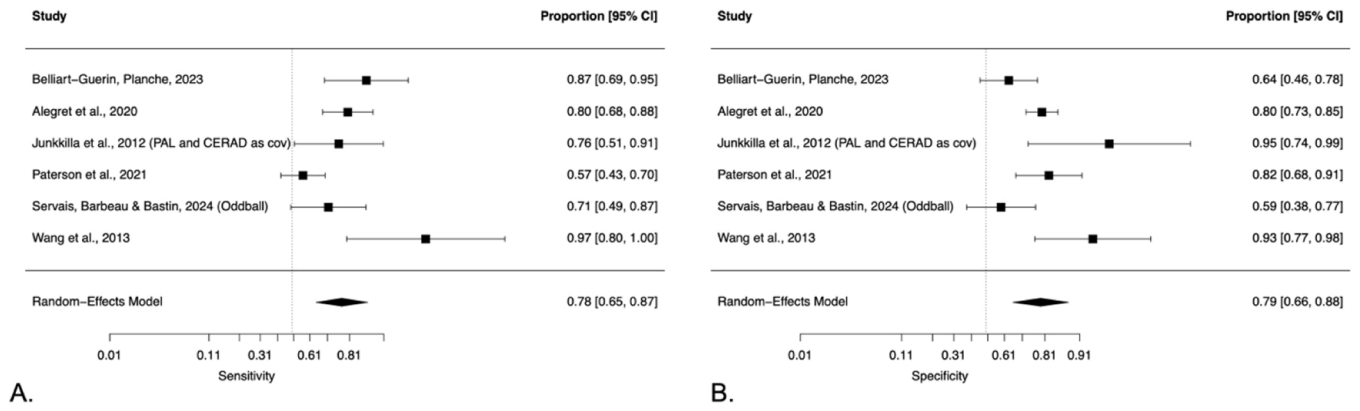

Fig. 2. A. Forest plots of diagnostic performance of digital episodic memory tests for detecting amnestic Mild Cognitive Impairment (aMCI). (A) Sensitivity estimates across included studies. (B) Specificity estimates across included studies. Squares represent study-level point estimates, with size proportional to study weight; horizontal lines indicate 95% confidence intervals (CI). Diamonds represent pooled estimates from random-effects models. The overall pooled sensitivity was 0.78 (95% CI: 065–0.87), and pooled specificity was 0.79 (95% CI: 0.66–0.88).

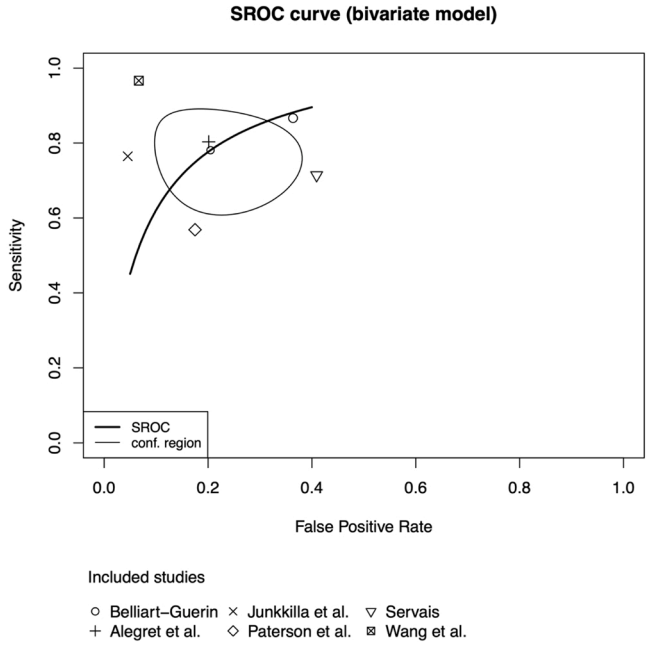

Fig. 3. Summary Receiver Operating Characteristic (SROC) curve based on a bivariate random-effects model. The SROC curve provides a global summary of diagnostic performance across studies evaluating digital episodic memory assessments for aMCI. Points represent study-level sensitivity and false positive rates. The solid black curve shows the fitted SROC, and the ellipse denotes the 95% confidence region surrounding the pooled estimate (open circle).

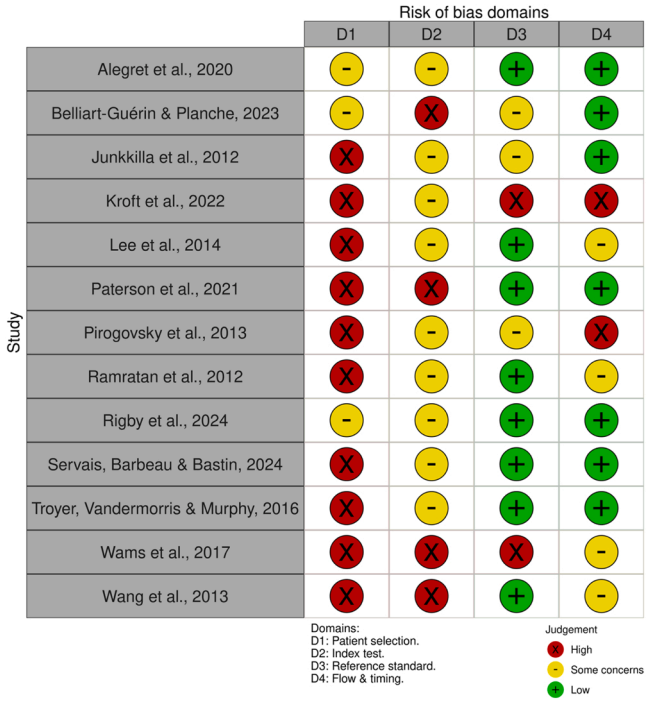

Fig. 4. Risk of bias assessment across included studies using the QUADAS-2 tool. Each study was evaluated across four domains: D1 – Patient selection, D2 – Index test, D3 – Reference standard, and D4 – Flow and timing. Judgements were categorized as Low risk (green, "+"), Some concerns (yellow, "–"), or High risk (red, "×"). Most studies demonstrated high risk of bias in at least one domain, particularly in patient selection and reference standard.

- Junkkila et al. (2012) assessed the diagnostic utility of the CANTAB Paired Associates Learning (PAL) task, a computerised visual memory test where participants learn and recall the locations of abstract patterns within boxes on a screen, with increasing levels of difficulty. The PAL total errors adjusted score effectively distinguished aMCI from healthy controls, achieving an AUC of 0.803 with an optimal cut-off of 42. While this was slightly lower than the CERAD Wordlist Learning Delayed Recall (AUC = 0.906), the PAL task still demonstrated strong discriminatory power as a standalone digital measure. When used in combination with the CERAD delayed recall score, classification accuracy improved to 84.5%, highlighting the complementary value of integrating computerised visuospatial tasks with traditional verbal memory measures in the detection of aMCI. To summarize, the data in this study supports the notion of the PAL total errors adjusted being the most valuable variable when attempting to distinguish between normal controls, aMCI patients and AD patients. Using ROC analysis, cut-off scores were calculated: a score of 42 best differentiated aMCI from healthy controls, while a score of 71 best differentiated aMCI from AD patients.
- Troyer et al. (2016) employed two computerised associative memory tasks to probe episodic memory in individuals with aMCI: a word–word association test and a face–name association test. In the word–word task, participants studied semantically unrelated word pairs and later completed a recognition test in which they had to identify whether pairs were intact, recombined, or entirely new. Similarly, in the face–name task, participants were presented with

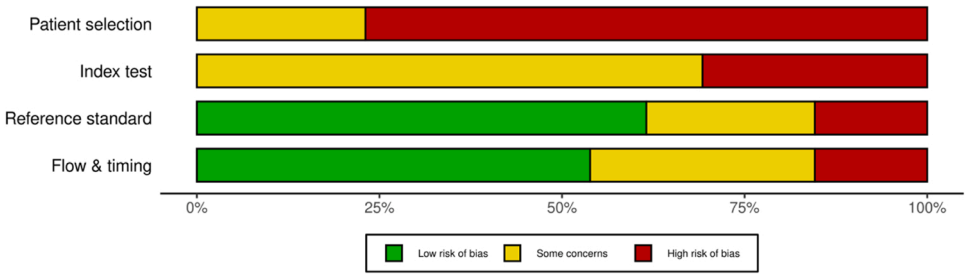

Fig. 5. Proportional summary of risk of bias across QUADAS-2 domains. This figure summarizes the distribution of risk of bias judgments across all included studies for each of the four QUADAS-2 domains: Patient selection, Index test, Reference standard, and Flow & timing. Bars represent the proportion of studies rated as having Low risk (green), Some concerns (yellow), or High risk (red) of bias per domain. Most studies were judged to have a high risk of bias in the Patient selection and Index test domains. Flow & timing showed a more balanced distribution, while Reference standard was most frequently rated as low risk.

faces accompanied by spoken names during the study phase, followed by a recognition test with intact, recombined, and novel pairings. The results showed that associative recognition accuracy, particularly for the word–word task, significantly differentiated individuals with aMCI from healthy controls, achieving 78% classification accuracy. When intra-individual variability (IIV) in response times was included as a feature, classification accuracy improved to 82%, suggesting that incorporating dynamic behavioural metrics may enhance the diagnostic sensitivity of associative memory paradigms.

- Servais et al. (2024) evaluated episodic memory through the Von Restorff memory task, designed to assess the benefit of distinctiveness on recall. Participants studied a list of 20 words, of which 4 were "distinctive" in either font, colour, or semantics. Following distraction, they performed a free recall test. The study found that while healthy controls benefited from distinctiveness, participants with aMCI recalled significantly fewer distinctive words, showing a reduced Von Restorff effect. In diagnostic analyses, memory-based indices derived from the task showed modest discriminative performance, characterised by high specificity but limited sensitivity, reflecting preserved recall of distinctive items in a subset of aMCI participants. This suggests that paradigms probing novelty effects, especially those leveraging salient cues, may detect early episodic memory impairments tied to medial temporal lobe dysfunction.
- Paterson et al. (2021) directly compared BHA, a self-administered digital tool, with the clinician-administered MoCA. Overall diagnostic performance was similar between the two (AUC = 0.76 for the BHA vs. 0.71 for the MoCA; p = .36), but the BHA demonstrated a higher corrected partial AUC (0.73 vs. 0.66), suggesting greater sensitivity in detecting milder impairments. The BHA model incorporated key memory-related subtests such as Face–Name Association and Spatial Working Memory which align closely with the mnemonic deficits typically seen in aMCI. Notably, the BHA classified a significantly smaller proportion of participants into the indeterminate probability category (56% vs. 70% with the MoCA; p = .004), which could help reduce unnecessary follow-ups or referrals.
- Alegret et al. (2020) developed FACEmemory®, a tablet-based associative memory task focused on face–name learning and recall, designed as a digital screening tool for early Alzheimer's disease and aMCI. Using a total score cut-off of 31.5, the test demonstrated good diagnostic accuracy in discriminating aMCI from cognitively unimpaired participants (AUC = 0.85), with balanced sensitivity (80.5%) and specificity (80%). The task targets associative binding processes known to be vulnerable in prodromal Alzheimer's disease and offers a brief, standardised, and clinically deployable digital alternative to traditional memory screening instruments.
- Among the included studies, Wang et al. (2013) demonstrated high diagnostic accuracy using a modified spatial-context memory task (SCMT), with an AUC of ~0.97, and sensitivity and specificity in the mid-90% range, strongly outperforming standard screening tools in distinguishing aMCI from healthy ageing. Their computerised spatial context paradigm presented participants with a city map showing multiple blocks and highlighted one with a flashing dot; participants subsequently learned to associate specific buildings with that block and were later cued to recall each building's location on the map. This task explicitly tests spatial context memory. This suggests that spatial context paradigms, grounded in hippocampal-dependent memory processes, may hold superior diagnostic value compared to more traditional episodic memory assessments. Highlighting these findings supports the argument that more targeted associative or relational memory tasks (such as spatial context or event-place association) could offer notable gains in sensitivity and specificity.

Our review builds on previous work (Germine et al., 2019; Thompson et al., 2023) by providing the first meta-analytic synthesis focused specifically on digital episodic memory tools. While the number of eligible studies remains limited, our pooled estimates offer a provisional benchmark for future validation. Study quality, assessed via QUADAS-2, often flagged selection bias and unclear reporting of reference standards, potentially inflating performance estimates. Furthermore, heterogeneity in task design, delivery, and lack of normative data, combined with small, non-diverse samples, limits generalisability.

Few of the included studies employed paradigms grounded in contemporary cognitive-neuroscience models in which episodic memory actively supports decision-making through retrieval of past experiences to guide future reasoning. (Shadlen and Shohamy, 2016). Instead, most tasks assessed episodic memory in isolation, without explicitly modelling its role in future-oriented or value-guided behaviour. The recent proposition on the function of memory in prospective inference (Buckner, 2010), the formation of associative links (Davachi, 2006; Rudy and Sutherland, 1989) and constructive episodic simulation (Addis et al., 2007) suggest that hippocampal activity can also have an active role in constructing de novo experiences in spatial and non-spatial contexts.

Newer tasks that assess higher-order cognitive domains, such as linking experiences across time, context, and features, processes thought to rely on memory and hippocampal-prefrontal circuitry, could hold translational potential into a clinical setting. By highlighting this neuronal mechanism and building insights into the neuronal computations underlying flexible behaviours that are difficult to study in animal models, but dominate human decision-making, we could identify subtler deficits that precede frank memory loss, a technique that could be especially valuable in recognising prodromal AD (Buzsáki, 2015; Jones et al., 2019). Future research should prioritise mechanistically informed paradigms such as associative binding, relational memory, and pattern separation tasks, which may better capture hippocampal dysfunction in early Alzheimer's disease.

In addition, the digital delivery of memory tasks enables collection of process-level data (e.g., response latencies, confidence ratings) that can be analysed using computational models. Such approaches could

quantify latent parameters of memory performance (Ratcliff and McKoon, 2022), offering a more fine-grained and mechanistic understanding of cognitive decline. Despite this potential, none of the included studies leveraged computational modelling, highlighting a key opportunity for future research.

Taken together, the present body of evidence on the diagnostic value of digital episodic memory tools for aMCI is currently characterised by substantial inconsistency. Reported sensitivities and specificities span a wide range across studies, reflecting differences in the cognitive construct being probed (e.g., associative recognition, pattern separation, spatial context, novelty detection), in the reference standards used to define aMCI, in sample composition and recruitment strategies, and in the analytic decisions used to derive cut-offs and diagnostic thresholds. Therefore, even when individual studies report promising performance, it remains difficult to determine whether comparable accuracy would be obtained in independent samples or in routine clinical settings. The exploratory pooled estimate reported above should therefore be read against this background of methodological heterogeneity, and as a description of where the literature currently stands rather than as a stable benchmark for digital episodic memory assessment.

4.2. Recommendations for future research

Several concrete steps would help resolve the inconsistencies identified above and move the field towards more comparable, clinically informative diagnostic accuracy work. First, future studies should adopt harmonised reporting of diagnostic accuracy metrics, including sensitivity, specificity, and AUC together with full confusion matrices, in line with STARD and emerging STARD-AI recommendations (Bossuyt et al., 2015). Second, the field would benefit from convergence on a small number of well-defined reference standards for aMCI, with explicit reporting of the criteria applied (e.g., Petersen, NIA-AA, Jak-Bondi actuarial criteria) and of the neuropsychological battery used to operationalise them. Third, cut-off thresholds for the digital index test should be pre-specified rather than derived post hoc in the same sample, with external validation in independent cohorts whenever possible. Studies should report the psychometric properties of the digital tool (test–retest reliability, learning effects, equivalence across hardware and language versions) alongside diagnostic performance, since these properties bear directly on real-world deployability. Also, the sample size should be planned with diagnostic accuracy estimation in mind, recognising that small case–control samples can yield unstable and optimistic performance estimates. Finally, mechanistically informed paradigms (e.g., associative binding, relational memory and high order cognitive functions that recruit episodic memory) should be evaluated alongside more traditional digital screeners in the same participants, to allow direct comparison and to clarify which features of episodic memory carry the greatest diagnostic signal in early aMCI.

5. Conclusions

This systematic review indicates that digital tools for measuring episodic memory yield moderate to high diagnostic accuracy for diagnosing aMCI, with AUC values generally above 0.80. Episodic memory tests can offer a finer-grained, more sensitive and specific alternative to traditional, multi-domain test batteries, as episodic memory is frequently the first domain to decline in Alzheimer's disease (Rizzolo et al., 2021).

Traditional neuropsychological assessment comprises multiple domain-specific tests, with diagnosis based on patterns of impairment and sparing. In contrast, brief cognitive screening instruments (e.g., MMSE, MoCA) rely on global composite scores and may be less sensitive to focal or early episodic memory deficits. In heterogeneous patient groups, reliance on global screening scores may therefore dilute sensitivity to early pathological change and increase false-positive classifications (Edmonds et al., 2015; Ganguli et al., 2011; Hobart et al., 2013; Kim et al., 2024a,b). Accordingly, there is growing interest in targeted, mechanistically informed tasks, particularly those probing hippocampal-dependent processes such as episodic and relational memory as complements to screening tools, with the potential to improve diagnostic precision and translational relevance.

While digital and computerised episodic memory tests show promise, at present they cannot replace clinical judgment and should not be used in isolation (Aslam et al., 2018). However, computerised methods allow for remote testing and scalability, providing bigger sample sizes taken from more diverse populations, as well as higher volumes of data. This could prove useful for screening and innovation in AD diagnosis. Multiple studies show provisional supportive evidence towards the feasibility and validity of computerised assessments in both supervised (Ding et al., 2022; Thompson et al., 2023) and unsupervised (Geddes et al., 2020) settings. Digital episodic-memory tasks may also serve as scalable triage tools, analogous to emerging blood-based biomarkers (e.g., plasma Aβ/tau), by identifying individuals who would benefit from comprehensive neuropsychological evaluation or biomarker testing rather than acting as stand-alone diagnostic instruments.

Further research should prioritise diverse samples, standardised task designs, and integration of tasks that assess higher-order cognitive functions. Developing tools sensitive to prodromal changes in Alzheimer's disease could improve screening accuracy and individualised care. In clinical settings, higher cut-off thresholds prioritise sensitivity to reduce false negatives and avoid missing true cases. In research contexts, lower thresholds can improve specificity and minimise false positives, thereby refining group-level comparisons.

Overall, digital episodic memory assessments may be particularly well-suited for early aMCI and early AD detection by assessing neurobiologically grounded mechanisms making them compelling targets for digital biomarkers.

CRediT authorship contribution statement

ARM: Concep., Meth., Formal Anal., Investig., Writing – Original Draft, DM: Data Curation, Writing – Review & Edit, CdS: Data Curation, Writing – Review & Edit, FM: Validation, Writing – Review & Edit, RH: Formal Anal., Writing – Review & Edit, EgW: Supervision; Meth., Writing – Review & Edit, RjD: Supervision, Writing – Review & Edit, IF: Project Administration, Investig., Writing – Review & Edit

Funding

This work was supported by a PhD Studentship from the Fundação para a Ciência e a Tecnologia (FCT, UI/BD/151509/2021), under the research units Comprehensive Health Research Centre (CHRC) and iNOVA4Health, NOVA Medical School, Universidade NOVA de Lisboa.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neubiorev.2026.106756.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon request.

References

Addis, D.R., Wong, A.T., Schacter, D.L., 2007. Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45, 1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016.

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., Phelps, C.H., 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7, 270–279. https://doi.org/10.1016/j.jalz.2011.03.008.

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., Phelps, C.H., 2013. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's disease. FOCUS 11, 96–106. https://doi.org/10.1176/appi.focus.11.1.96.

Alegret, M., Munoz, N., Roberto, N., Rentz, D.M., Valero, S., Gil, S., Marquie, M., Hernandez, I., Riveros, C., Sanabria, A., Perez-Cordon, A., Espinosa, A., Ortega, G., Mauleon, A., Abdelnour, C., Rosende-Roca, M., Papp, K.V., Orellana, A., Benaque, A., Tarraga, L., Ruiz, A., Boada, M., 2020. A computerized version of the short form of the face-name associative memory exam (FACEmemory) for the early detection of Alzheimer's disease. Alzheimers Res. Ther. 12, 25. https://doi.org/10.1186/s13195-020-00594-6.

Alves, L., Cardoso, S., Silva, D., Mendes, T., Maróco, J., Nogueira, J., Lima, M., Tábuas-Pereira, M., Baldeiras, I., Santana, I., Mendonça, A. de, Guerreiro, M., 2021. Neuropsychological profile of amyloid-positive versus amyloid-negative amnestic Mild Cognitive Impairment. J. Neuropsychol. 15, 41–52. https://doi.org/10.1111/jnp.12218.

Aslam, R.W., Bates, V., Dundar, Y., Hounsome, J., Richardson, M., Krishan, A., Dickson, R., Boland, A., Fisher, J., Robinson, L., Sikdar, S., 2018. A systematic review of the diagnostic accuracy of automated tests for cognitive impairment. Int. J. Geriatr. Psychiatry 33, 561–575. https://doi.org/10.1002/gps.4852.

Babb, S.J., Johnson, R.M., 2010. Object, spatial, and temporal memory: a behavioral analysis of visual scenes using a what, where, and when paradigm. Curr. Psychol. Lett. Behav. Brain Cogn. 26.

Belliart-Guérin, G., Planche, V., 2023. Mnemonic discrimination performance in a memory clinic: a pilot study. J. Alzheimers Dis. 94, 1527–1534. https://doi.org/10.3233/jad-230221.

Bossuyt, P.M., Deeks, J.J., Leeflang, M.M., Takwoingi, Y., Flemyng, E., 2023. Evaluating medical tests: introducing the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Cochrane Database Syst. Rev. 2023. https://doi.org/10.1002/14651858.ed000163.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L., et al., 2015. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351, h5527. https://doi.org/10.1136/bmj.h5527.

Buckner, R.L., 2010. The role of the Hippocampus in prediction and imagination. Annu. Rev. Psychol. 61, 27–48. https://doi.org/10.1146/annurev.psych.60.110707.163508.

Buzsáki, G., 2015. Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. Hippocampus 25, 1073–1188. https://doi.org/10.1002/hipo.22488.

Chi, S.Y., Rabin, L.A., Aronov, A., Fogel, J., Kapoor, A., Wang, C., 2014 Nov. Differential focal and nonfocal prospective memory accuracy in a demographically diverse group of nondemented community-dwelling older adults. J Int Neuropsychol Soc. 20 (10), 1015–1027. https://doi.org/10.1017/s1355617714000964. Epub 2014 Nov 17. PMID: 25401793; PMCID: PMC4624305.

Cirulli, F., Easton, A., 2022 Nov. Placing behaviour at the forefront of brain science. Neurosci Biobehav Rev. 142, 104861. https://doi.org/10.1016/j.neubiorev.2022.104861. Epub 2022 Sep 6. PMID: 36075343.

Cogstate: Simplifying the Measurement of Cognition [WWW Document], n.d. URL ⟨https://www.cogstate.com/?gad_source=1&gclid=CjwKCAiAxKy5BhBbEiwAYiW–zeZArl94Uu1CzN8T9-EyC4_RcaqgizAvyDGvi6eRQlPEGOkC86QXBoCJP8QAvD_BwE&cn-reloaded=1⟩ (accessed 11.6.24).

Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. Mot. Syst. / Neurobiol. Behav. 16, 693–700. https://doi.org/10.1016/j.conb.2006.10.012.

Deeks, J.J., Macaskill, P., Irwig, L., 2005. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J. Clin. Epidemiol. 58, 882–893. https://doi.org/10.1016/j.jclinepi.2005.01.016.

Delhaye, E., Mechanic-Hamilton, D., Saad, L., Das, S.R., Wisse, L.E.M., Yushkevich, P.A., Wolk, D.A., Bastin, C., 2019b. Associative memory for conceptually unitized word pairs in mild cognitive impairment is related to the volume of the perirhinal cortex. Hippocampus 29, 630–638. https://doi.org/10.1002/hipo.23063.

Delhaye, E., Mechanic-Hamilton, D., Saad, L., Das, S.R., Wisse, L.E.M., Yushkevich, P.A., Wolk, D.A., Bastin, C., 2019a. Associative memory for conceptually unitized word pairs in mild cognitive impairment is related to the volume of the perirhinal cortex. Hippocampus 29, 630–638. https://doi.org/10.1002/hipo.23063.

Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed, 2013., Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed. American Psychiatric Publishing, Inc., Arlington, VA, US. https://doi.org/10.1176/appi.books.9780890425596.

Ding, Z., Lee, T., Chan, A.S., 2022. Digital cognitive biomarker for mild cognitive impairments and dementia: a systematic review. J. Clin. Med. 11, 4191. https://doi.org/10.3390/jcm11144191.

Doebler, P., Holling, H., 2015. Meta-analysis of diagnostic accuracy and ROC curves with covariate adjusted semiparametric mixtures. Psychometrika 80, 1084–1104. https://doi.org/10.1007/s11336-014-9430-0.

Edmonds, E.C., Delano-Wood, L., Clark, L.R., Jak, A.J., Nation, D.A., McDonald, C.R., Libon, D.J., Au, R., Galasko, D., Salmon, D.P., Bondi, M.W., 2015. Susceptibility of the conventional criteria for MCI to false positive diagnostic errors. Alzheimers Dement. J. Alzheimers Assoc. 11, 415–424. https://doi.org/10.1016/j.jalz.2014.03.005.

Égerházi, A., Berecz, R., Bartók, E., Degrell, I., 2007. Automated neuropsychological test battery (CANTAB) in mild cognitive impairment and in Alzheimer's disease. Prog. Neuropsychopharmacol. Biol. Psychiatry 31, 746–751. https://doi.org/10.1016/j.pnpbp.2007.01.011.

Eichenbaum, H., 2017. The role of the hippocampus in navigation is memory. J. Neurophysiol. 117, 1785–1796. https://doi.org/10.1152/jn.00005.2017.

Ganguli, M., Snitz, B.E., Saxton, J.A., Chang, C.-C.H., Lee, C.-W., Bilt, J.V., Hughes, T.F., Loewenstein, D.A., Unverzagt, F.W., Petersen, R.C., 2011. Outcomes of mild cognitive impairment depend on definition: a population study. Arch. Neurol. 68, 761–767. https://doi.org/10.1001/archneurol.2011.101.

Geddes, M.R., O'Connell, M.E., Fisk, J.D., Gauthier, S., Camicioli, R., Ismail, Z., COVID-19, for the A.S. of C.T.F. on D.C.B.P. for, 2020. Remote cognitive and behavioral assessment: Report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID-19. Alzheimers Dement. Diagn. Assess. Dis. Monit. 12, e12111. https://doi.org/10.1002/dad2.12111.

Germine, L., Reinecke, K., Chaytor, N.S., 2019. Digital neuropsychology: challenges and opportunities at the intersection of science and software. Clin. Neuropsychol. 33, 271–286. https://doi.org/10.1080/13854046.2018.1535662.

Gershon, R.C., Cella, D., Fox, N.A., Havlik, R.J., Hendrie, H.C., Wagster, M.V., 2010. Assessment of neurological and behavioural function: the NIH Toolbox. Lancet Neurol. 9, 138–139. https://doi.org/10.1016/S1474-4422(09)70335-7.

Guillery-Girard, B., Martins, S., Deshayes, S., Hertz-Pannier, L., Chiron, C., Jambaqué-Aubourg, I., Landeau, B., Clochon, P., Chetelat, G., Eustache, F., 2013. Developmental trajectories of associative memory from childhood to adulthood: a behavioral and neuroimaging study. Front. Behav. Neurosci. 7. https://doi.org/10.3389/fnbeh.2013.00126.

Gustavson, D.E., Elman, J.A., Sanderson-Cimino, M., Franz, C.E., Panizzon, M.S., Jak, A.J., Reynolds, C.A., Neale, M.C., Lyons, M.J., Kremen, W.S., 2020. Extensive memory testing improves prediction of progression to MCI in late middle age. Alzheimers Dement. Diagn. Assess. Dis. Monit. 12, e12004. https://doi.org/10.1002/dad2.12004.

Hansson, O., 2021. Biomarkers for neurodegenerative diseases. Nat. Med. 27, 954–963. https://doi.org/10.1038/s41591-021-01382-x.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327, 557–560. https://doi.org/10.1136/bmj.327.7414.557.

Hobart, J., Cano, S., Posner, H., Selnes, O., Stern, Y., Thomas, R., Zajicek, J., Initiative, A.D.N., 2013. Putting the Alzheimer's cognitive test to the test II: rasch measurement theory. Alzheimers Dement 9, S10–S20. https://doi.org/10.1016/j.jalz.2012.08.006.

Jack, C.R., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, E.G., Kokmen, E., 2000. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. Neurology 55, 484–490. https://doi.org/10.1212/WNL.55.4.484.

Jones, E.A., Gillespie, A.K., Yoon, S.Y., Frank, L.M., Huang, Y., 2019. Early hippocampal sharp-wave ripple deficits predict later learning and memory impairments in an Alzheimer's disease mouse model. Cell Rep. 29, 2123–2133.e4. https://doi.org/10.1016/j.celrep.2019.10.056.

Junkkila, J., Oja, S., Laine, M., Karrasch, M., 2012. Applicability of the CANTAB-PAL computerized memory test in identifying amnestic mild cognitive impairment and Alzheimer's disease. Dement. Geriatr. Cogn. Disord. 34, 83–89. https://doi.org/10.1159/000342116.

Kim, H.-G., Kim, Y., Edmonds, E.C., Bondi, M.W., 2024. Comparison of cognitive impairment diagnosis criteria in clinical settings: conventional vs. neuropsychological. Alpha Psychiatry 25, 212–219. https://doi.org/10.5152/alphapsychiatry.2024.231448.

Kim, H., Lee, S., Levine, A., Huber, B., Andrews, H., Kerner, N.A., Cohen, D., Carlson, S., Bell, S.A., Rivera, A.M., Gordon, M.L., Simoes, S., Devanand, D.P., Brickman, A.M., Schneider, L.S., Harvey, P.D., Goldberg, T.E., 2024. Novel measures of cognition and function for the AD spectrum in the Novel Measures for Alzheimer's Disease Prevention Trials (NoMAD) project: psychometric properties, convergent validation, and contrasts with established measures. Alzheimers Dement. J. Alzheimers Assoc. 20, 5089–5101. https://doi.org/10.1002/alz.13757.

Knopman, D.S., Parisi, J.E., Salviati, A., Floriach-Robert, M., Boeve, B.F., Ivnik, R.J., Smith, G.E., Dickson, D.W., Johnson, K.A., Petersen, L.E., McDonald, W.C., Braak, H., Petersen, R.C., 2003 Nov. Neuropathology of cognitively normal elderly. J Neuropathol Exp Neurol. 62 (11), 1087–1095. https://doi.org/10.1093/jnen/62.11.1087. PMID: 14656067.

Kourtis, L.C., Regele, O.B., Wright, J.M., Jones, G.B., 2019. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. Npj Digit. Med. 2, 9. https://doi.org/10.1038/s41746-019-0084-2.

Kroft, D., Abo, A.M., Meek, B., Bolster, B., Modirrousta, M.O., 2022. Determining affected memory domains in patients with amnestic mild cognitive impairment using computerized and interactive tests. Appl. Neuropsychol. Adult 29, 1530–1535. https://doi.org/10.1080/23279095.2021.1896518. ⟨https://orcid.org/0000-0003-3247-326X M.M.A⟩.

LaPlume, A.A., Paterson, T.S.E., Gardner, S., Stokes, K.A., Freedman, M., Levine, B., Troyer, A.K., Anderson, N.D., 2021. Interindividual and intraindividual variability in amnestic mild cognitive impairment (aMCI) measured with an online cognitive assessment. J. Clin. Exp. Neuropsychol. 43, 796–812. https://doi.org/10.1080/13803395.2021.1982867.

Lawrence, E., Vegvari, C., Ower, A., Hadjichrysanthou, C., Wolf, F.D., Anderson, R.M., 2017. A systematic review of longitudinal studies which measure Alzheimer's disease biomarkers. J. Alzheimers Dis. Prepr. 1–21. https://doi.org/10.3233/jad-170261.

Lee, J.-Y., Kho, S., Yoo, H.B., Park, S., Choi, J.-S., Kwon, J.S., Cha, K.R., Jung, H.-Y., 2014. Spatial memory impairments in amnestic mild cognitive impairment in a virtual radial arm maze. Neuropsychiatr. Dis. Treat. 653. https://doi.org/10.2147/NDT.S58185.

Li, X., Jia, S., Zhou, Z., Jin, Y., Zhang, X., Hou, C., Zheng, W., Rong, P., Jiao, J., 2018. The role of the Montreal Cognitive Assessment (MoCA) and its memory tasks for detecting mild cognitive impairment. Neurol. Sci. 39, 1029–1034. https://doi.org/10.1007/s10072-018-3319-0.

Lipsey, M.W., 2001. Practical Meta-analysis. Sage Publications, Thousand Oaks, Calif.

Malek-Ahmadi, M., Nikkhahmanesh, N., 2024. Meta-analysis of Montreal cognitive assessment diagnostic accuracy in amnestic mild cognitive impairment. Front. Psychol. 15, 1369766. https://doi.org/10.3389/fpsyg.2024.1369766.

McInnes, M.D.F., Moher, D., Thombs, B.D., McGrath, T.A., Bossuyt, P.M., and the PRISMA-DTA Group, 2018. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. JAMA 319, 388–396. https://doi.org/10.1001/jama.2017.19163.

Ossenkoppele, R., Binette, A.P., Groot, C., Smith, R., Strandberg, O., Palmqvist, S., Stomrud, E., Tideman, P., Ohlsson, T., Jögi, J., Johnson, K., Sperling, R., Dore, V., Masters, C.L., Rowe, C., Visser, D., Berckel, B.N.M. van, Flier, W.M. van der, Baker, S., Jagust, W.J., Wiste, H.J., Petersen, R.C., Jack, C.R., Hansson, O., 2022. Amyloid and tau PET-positive cognitively unimpaired individuals are at high risk for future cognitive decline. Nat. Med. 28, 2381–2387. https://doi.org/10.1038/s41591-022-02049-x.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan—a web and mobile app for systematic reviews. Syst. Rev. 5, 210. https://doi.org/10.1186/s13643-016-0384-4.

Paterson, T.S.E., Sivajohan, B., Gardner, S., Binns, M.A., Stokes, K.A., Freedman, M., Levine, B., Troyer, A.K., 2022. Accuracy of a self-administered online cognitive assessment in detecting amnestic mild cognitive impairment. J. Gerontol. Ser. B 77, 341–350. https://doi.org/10.1093/geronb/gbab097.

Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. J. Intern. Med. 256, 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x.

Petersen, R.C., 2016. Mild cognitive impairment. Contin. Lifelong Learn. Neurol. 22, 404–418. https://doi.org/10.1212/con.0000000000000313.

Petersen, R.C., Doody, R., Kurz, A., et al., 2001. Current concepts in mild cognitive impairment. Arch. Neurol. 58 (12), 1985–1992. https://doi.org/10.1001/archneur.58.12.1985.

Petersen, R.C, Negash, S., 2008 Jan. Mild cognitive impairment: an overview. CNS Spectr. 13 (1), 45–53. https://doi.org/10.1017/s1092852900016151. PMID: 18204414.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive Impairment: clinical characterization and outcome. Arch. Neurol. 56 (3), 303–308. https://doi.org/10.1001/archneur.56.3.303. Erratum in: Arch Neurol 1999 Jun;56(6):760. PMID: 10190820.

Pirogovsky, E., Holden, H.M., Jenkins, C., Peavy, G.M., Salmon, D.P., Galasko, D.R., Gilbert, P.E., 2013. Temporal sequence learning in healthy aging and amnestic mild cognitive impairment. Exp. Aging Res. 39, 371–381. https://doi.org/10.1080/0361073x.2013.808122.

Ramratan, W.S., Rabin, L.A., Wang, C., Zimmerman, M.E., Katz, M.J., Lipton, R.B., Buschke, H., 2012. Level of recall, retrieval speed, and variability on the Cued-Recall Retrieval Speed Task (CRRST) in individuals with amnestic mild cognitive impairment. J. Int. Neuropsychol. Soc. 18, 260–268. https://doi.org/10.1017/s1355617711001664.

Ratcliff, R., McKoon, G., 2022. Can neuropsychological testing be improved with model-based approaches? Trends Cogn. Sci. 26, 899–901. https://doi.org/10.1016/j.tics.2022.08.015.

Rigby, T., Gregoire, A.M., Reader, J., Kahsay, Y., Fisher, J., Kairys, A., Bhaumik, A.K., Rahman-Filipiak, A., Maher, A.C., Hampstead, B.M., Heidebrink, J.L., Kavcic, V., Giordani, B., 2024. Identification of amnestic mild cognitive impairment among Black and White community-dwelling older adults using NIH Toolbox Cognition tablet battery. J. Int. Neuropsychol. Soc. 30, 689–696. https://doi.org/10.1017/S1355617724000213.

Rizzolo, L., Narbutas, J., Van Egroo, M., Chylinski, D., Besson, G., Baillet, M., Ali Bahri, M., Salmon, E., Maquet, P., Vandewalle, G., Bastin, C., Collette, F., 2021. Relationship between brain AD biomarkers and episodic memory performance in healthy aging. Brain Cogn. 148, 105680. https://doi.org/10.1016/j.bandc.2020.105680.

Rudy, J.W., Sutherland, R.J., 1989. The hippocampal formation is necessary for rats to learn and remember configural discriminations. Behav. Brain Res. 34, 97–109. https://doi.org/10.1016/S0166-4328(89)80093-2.

Salameh, J.-P., Bossuyt, P.M., McGrath, T.A., Thombs, B.D., Hyde, C.J., Macaskill, P., Deeks, J.J., Leeflang, M., Korevaar, D.A., Whiting, P., Takwoingi, Y., Reitsma, J.B., Cohen, J.F., Frank, R.A., Hunt, H.A., Hooft, L., Rutjes, A.W.S., Willis, B.H., Gatsonis, C., Levis, B., Moher, D., McInnes, M.D.F., 2020. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ m2632. https://doi.org/10.1136/bmj.m2632.

Saxton, J., Morrow, L., Eschman, A., Archer, G., Luther, J., Zuccolotto, A., 2009. Computer assessment of mild cognitive impairment. Postgrad. Med. 121, 177–185. https://doi.org/10.3810/pgm.2009.03.1990.

Schwarzer, G., Carpenter, J.R., Rücker, G., 2015. Meta-Analysis with R, Use R! Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-21416-0.

Servais, A., Barbeau, E.J., Bastin, C., 2024. Contextual novelty detection and novelty-related memory enhancement in amnestic mild cognitive impairment. Cortex 172, 72–85. https://doi.org/10.1016/j.cortex.2023.12.005.

Shadlen, M.N., Shohamy, D., 2016. Decision making and sequential sampling from memory. Neuron 90, 927–939. https://doi.org/10.1016/j.neuron.2016.04.036.

Shohamy, D., Daw, N.D., 2015. Integrating memories to guide decisions. Curr. Opin. Behav. Sci. 5, 85–90. https://doi.org/10.1016/j.cobeha.2015.08.010.

Summers, M.J., Saunders, N.L.J., 2012. Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment. Neuropsychology 26, 498–508. https://doi.org/10.1037/a0028576.

Thompson, L.I., Kunicki, Z.J., Emrani, S., Strenger, J., Vito, A.N.D., Britton, K.J., Dion, C., Harrington, K.D., Roque, N., Salloway, S., Sliwinski, M.J., Correia, S., Jones, R.N., 2023. 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. https://doi.org/10.1002/dad2.12500.

Troyer, A.K., Vandermorris, S., Murphy, K.J., 2016. Intraindividual variability in performance on associative memory tasks is elevated in amnestic mild cognitive impairment. Neuropsychologia 90, 110–116. https://doi.org/10.1016/j.neuropsychologia.2016.06.011.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor Package. J. Stat. Softw. 36, 1–48. https://doi.org/10.18637/jss.v036.i03.

Wams, E.J., Wilcock, G.K., Foster, R.G., Wulff, K., 2017. Sleep-Wake patterns and cognition of older adults with amnestic mild cognitive impairment (aMCI): a comparison with cognitively healthy adults and moderate Alzheimer's disease patients. Curr. Alzheimer Res. 14. https://doi.org/10.2174/1567205014666170523095634.

Wang, H.-M., Yang, C.-M., Kuo, W.-C., Huang, C.-C., Kuo, H.-C., 2013. Use of a modified spatial-context memory test to detect amnestic mild cognitive impairment. PLoS One 8, e57030. https://doi.org/10.1371/journal.pone.0057030.

Whiting, P.F., Rutjes, A.W.S., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., Bossuyt, P.M.M., QUADAS-2 Group, 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155, 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., De Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., Van Duijn, C., Visser, P., Petersen, R.C., 2004. Mild cognitive impairment – beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. J. Intern. Med. 256, 240–246. https://doi.org/10.1111/j.1365-2796.2004.01380.x.

Wolk, D.A., Price, J.C., Saxton, J.A., Snitz, B.E., James, J.A., Lopez, O.L., Aizenstein, H.J., Cohen, A.D., Weissfeld, L.A., Mathis, C.A., Klunk, W.E., De-Kosky, S.T., DeKoskym, S.T., 2009. Amyloid imaging in mild cognitive impairment subtypes. Ann. Neurol. 65, 557–568. https://doi.org/10.1002/ana.21598.