Scientific Review
UDC 159.9
https://doi.org/10.21702/rpj.2026.3.2

Meta-analysis of Diagnostic Accuracy of Traditional vs. Technology-based Cognitive Function Assessments

Nazyar Khamenehei*, Lyudmila V. Tokarskaya

Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, Russian Federation

*Corresponding author: Nazyarkh@gmail.com

Abstract

Introduction. Cognitive function assessment plays a vital role in diagnosing and monitoring neurocognitive disorders. While traditional tools like the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) are widely used, they face limitations in sensitivity, ecological validity, and adaptability. With advancements in technology, alternative methods—such as Virtual Reality (VR), Artificial Intelligence (AI), and machine learning—have emerged, offering potential improvements in diagnostic accuracy and user engagement. This meta-analysis compares the effectiveness and usability of traditional and technological cognitive assessments.

Methods. A systematic review of 42 studies from PubMed, Scopus, and Web of Science was conducted, focusing on diagnostic accuracy, usability, engagement, and reliability. Statistical analyses included sensitivity, specificity, effect sizes.

Results. Results show that technological methods—particularly AI-based models, VR assessments, gamified tasks, and wearable systems—achieve higher diagnostic accuracy (AUC=0.84–0.91) than traditional methods (AUC=0.60–0.80), especially in early detection of mild cognitive impairment (MCI) and dementia. VR and gamified tests also showed increased engagement and ecological validity. However, digital approaches present challenges in standardization, data privacy, and accessibility for older adults.

Discussion. Despite lower sensitivity, traditional methods remain prevalent due to their low

cost, ease of use, and clinical acceptance. A hybrid model combining both approaches may enhance diagnostic outcomes. While technology-integrated solutions hold great promise for cognitive evaluation, further research is needed to validate AI models in diverse populations and address barriers to accessibility and integration into clinical practice.

Keywords

cognitive assessment, virtual reality, artificial intelligence, serious games, wearable sensors, traditional tests, ecological validity, cognitive flexibility

For citation

Khamenehei, N., & Tokarskaya, L. V. (2026). Meta-analysis of diagnostic accuracy of traditional vs. technology-based cognitive function assessments. *Russian Psychological Journal*, *23*(3), 32–65. https://doi.org/10.21702/rpj.2026.3.2

Introduction

Assessing cognitive function plays a fundamental role in the diagnosis, differential diagnosis, and longitudinal monitoring of neurocognitive conditions, including dementia, Alzheimer's disease, and mild cognitive impairment (MCI). Traditional standardized instruments—most prominently the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), and related paper-and-pencil measures—are widely used to evaluate memory, attention, language, processing speed, and executive functioning, and they remain integral to routine screening (Barboza E Barbosa & Fichman, 2019; Livingston et al., 2020; Alonso-Quiñones et al., 2021; Kourtis et al., 2019; Dobie et al., 2025; Kurbanova et al., 2020). Their clinical importance lies not only in case finding but also in establishing baselines against which cognitive trajectories can be tracked over time and care individualized (Winblad et al., 2016; Silverberg et al., 2011; Hildebrandt, 2020). At the same time, early cognitive decline is often subtle and heterogeneous, demanding instruments with high sensitivity and specificity (Albert et al., 2011). Long-standing critiques highlight cultural and educational bias and limited ecological validity in conventional tests, which frequently fail to approximate real-world cognitive demands and can therefore constrain generalizability and interpretability (Dawson & Marcotte, 2017; Manly, 2008). These limitations have motivated efforts to enhance both the accuracy and efficiency of cognitive evaluation.

Cognitive assessment is equally crucial in pediatric contexts, where it supports the identification of developmental delays, the diagnosis of learning disabilities, and the design of tailored educational interventions. Established batteries such as the Wechsler Intelligence Scale for Children and the Stanford–Binet Intelligence Scales remain central for appraising reasoning, memory, and processing speed. In parallel, digital tools have begun to enter practice.

Notably, the QbTest—approved for clinical use within the NHS in England and Wales—provides objective indices of inattention, impulsivity, and hyperactivity to support Attention-Deficit/Hyperactivity Disorder (ADHD) diagnosis, while instruments like the Cognitive Assessment of Young Children (CAYC) offer structured, play-based approaches for children from two months to five years (Urbina et al., 2014; Bellato et al., 2024). These developments aim to increase precision, efficiency, and accessibility in pediatric assessment without displacing the interpretive strengths of traditional measures.

Converging advances in technology further expand the assessment toolkit across the lifespan. Virtual Reality (VR) enables ecologically valid simulations of everyday cognitive demands—such as navigation, multitasking, and goal management—while preserving experimental control over stimuli and contingencies. In turn, Artificial Intelligence (AI) and machine-learning methods can leverage high-resolution behavioral traces generated in these environments (e.g., timing, kinematics) to improve diagnostic discrimination, including between cognitively healthy individuals and those with MCI or early-stage Alzheimer's disease. Empirical studies report promising validity, reliability, and efficiency for VR-based platforms; they also underscore the advantages of precise stimulus control and detailed response logging relative to conventional formats. Collectively, these advanced technological systems for neuropsychological assessment (ATSNA)—encompassing VR, AI, Serious Games, and wearable sensors—hold considerable potential to deliver comprehensive, engaging, and scalable evaluations with stronger ecological validity than legacy paper-based tools (Parsons, 2011; Pedroli et al., 2015; Jo et al., 2019; Liu et al., 2019; Kourtis et al., 2019).

In sum, while traditional instruments remain indispensable anchors of clinical decision-making, the field is undergoing a measured shift toward technology-enhanced approaches designed to address recognized constraints—low sensitivity to incipient decline, cultural and educational bias, and limited ecological validity—without sacrificing interpretability or feasibility.

Historical Evolution of Cognitive Assessment Methods

Over the past century, cognitive assessment has progressed from rudimentary paper-and-pencil instruments to increasingly standardized and psychometrically robust tools, mirroring advances in neuropsychology and measurement science (C. Li et al., 2024). The earliest efforts emphasized broad appraisal of intellectual ability and problem-solving, most notably the Binet–Simon Intelligence Test (1905), which sought to identify school-aged children in need of specialized educational support by sampling attention, memory, and verbal skills (Eling, 2019). Mid-century work by Wechsler consolidated and extended this tradition: the Wechsler–Bellevue Intelligence Scale and its successor, the Wechsler Adult Intelligence Scale, formalized separate verbal and performance indices, establishing a framework that enabled more nuanced profiling of cognitive strengths and weaknesses and that profoundly influenced modern neuropsychological testing (Boake, 2002).

From these foundations, domain-specific clinical screens and performance measures were introduced to target particular cognitive systems. The Mini-Mental State Examination

(MMSE) became a ubiquitous brief screen from 1975 onward (Folstein et al., 1975), and the Montreal Cognitive Assessment (MoCA) improved sensitivity to mild cognitive impairment while remaining sensitive to educational background (Nasreddine et al., 2005). Complementary performance tests expanded coverage beyond global cognition: the Trail Making Test (TMT) indexed set-shifting and processing speed but could be affected by motor demands and cross-cultural factors (Fernández & Marcopulos, 2008), whereas the Digit Symbol Substitution Test (DSST) provided an efficient probe of processing speed and working memory yet could be confounded by age-related motor slowing (Jaeger, 2018).

Despite their practical value, limitations of traditional methods are well documented. Cultural and educational biases can compromise fairness and interpretability across diverse populations (Manly, 2008). Low ecological validity—stemming from decontextualized, clinic-based tasks—reduces correspondence with everyday functioning (Chaytor & Schmitter-Edgecombe, 2003). Moreover, monotonous task formats may diminish motivation and engagement, with consequences for performance (Khaleghi et al., 2021), and fixed item structures restrict adaptability to individual ability levels (Zorowitz & Niv, 2023). Collectively, these challenges have catalyzed the search for more dynamic, engaging, and ecologically valid approaches to cognitive assessment, setting the stage for contemporary innovations while preserving the foundational insights of traditional testing.

Emergence of Technological Cognitive Assessment Methods

Advancements in digital technology have transformed cognitive assessment by introducing interactive, immersive, and adaptive tools that enhance accuracy, engagement, and ecological validity (Horan et al., 2024; Chicchi Giglioli, I. A. M. ,2020). The key innovations include digital platforms, virtual reality (VR), mobile apps, wearables, and machine learning models.

Virtual Reality (VR)

Virtual Reality (VR) has revolutionized cognitive assessment by creating immersive 3D environments that simulate real-world tasks, thereby enhancing ecological validity and cognitive engagement (Bhargava et al., 2024; Romero-Ayuso et al., 2024; Faria et al., 2023). Virtual Reality (VR) enables dynamic interaction, real-time feedback, and complex cognitive tasks that closely mimic daily activities (Sokołowska, 2023; Borgnis et al., 2022; Buele et al., 2023).

One of the key advantages of VR in cognitive assessment is its *enhanced ecological validity*. VR has significantly improved the realism of cognitive assessments by providing simulations that closely mirror real-world tasks, thereby facilitating the evaluation of executive functions such as planning, problem-solving, and working memory (Borgnis et al., 2022; Kirkham et al., 2024). Additionally, VR has been shown to *increase engagement and motivation* among users by providing immersive experiences that reduce test anxiety and enhance motivation (Pratviel et al., 2024; Mouatt et al., 2020; Wilczyńska et al., 2025). VR also supports dynamic and adaptive assessments, enabling personalized cognitive tasks with adaptive difficulty levels, thus improving accuracy and sensitivity (Lui, Lasse et al., 2025; Zhou et al., 2023).

Despite these advantages, VR-based cognitive assessments face certain limitations and challenges. Technological dependency is one such challenge, as implementing VR assessments requires specific hardware and a certain level of digital literacy, which can be problematic for older adults (Vercruyssen et al., 2023; Seifert & Schlomann, 2021; White, 2023). Additionally, motion sickness and discomfort are common issues in VR environments, potentially affecting user performance. These symptoms often arise from a mismatch between visual inputs and physical sensations, leading to discomfort (Chang et al., 2020).

Artificial Intelligence (AI) and Machine Learning (ML)

Artificial Intelligence (AI) and Machine Learning (ML) have become pivotal in analyzing extensive datasets to identify patterns of cognitive decline, facilitating early diagnosis and personalized cognitive assessments (Kale et al., 2024; W. Y. Tan et al., 2023). Advanced algorithms, such as neural networks and gradient boosting, have been employed to predict cognitive impairment with notable accuracy. These models analyze complex datasets, including neuroimaging and clinical assessments, to identify patterns indicative of cognitive decline (Musto et al., 2024; H. Li & Fan, 2019; Grammenos et al., 2024).

AI models have demonstrated high diagnostic accuracy and sensitivity in detecting mild cognitive impairment (MCI) and dementia compared to traditional assessment methods. Recent studies have highlighted the efficacy of AI in early diagnosis and differentiation of cognitive disorders (Hao et al., 2024; Obuchi et al., 2024; Xue et al., 2024). AI-based assessments offer scalability and automation, providing automated scoring that reduces examiner bias and improves scalability for telemedicine applications (Lesport et al., 2024; Migal & Nazarov, 2024). Furthermore, AI models excel in predictive analytics and early detection, as they can identify subtle cognitive changes earlier than traditional tests, enabling proactive interventions (Hao et al., 2024; Shankar et al., 2025).

Despite these advantages, AI methodologies present certain limitations and challenges. Data privacy and security are significant concerns, as AI models require large datasets, raising ethical issues regarding the safeguarding of sensitive information (OVIC, 2021; ISC2, 2024). Additionally, algorithmic bias and generalizability remain critical challenges, as AI models may exhibit bias due to imbalanced training data, which can affect their applicability and fairness across diverse populations (Yang et al., 2023).

Serious Games

Serious Games are interactive digital games developed for purposes beyond entertainment—most notably cognitive assessment, training, and rehabilitation—by embedding validated cognitive tasks within engaging, gamified experiences (Guzmán et al., 2024; Thongthip et al., 2024; Shahmoradi et al., 2022; Abd-Alrazaq et al., 2023). Typical designs leverage core game mechanics—rewards, progressive challenges, and adaptive difficulty—to sustain motivation while personalizing task demands to an individual's performance profile. Delivery across mobile phones, tablets, desktop systems, and immersive virtual reality extends reach and

usability in both clinical and community contexts (Hammady & Arnab, 2022; Hendrix et al., 2019; Streicher & Smeddinck, 2016).

As assessment tools, Serious Games can target memory, attention, executive functions, and problem-solving within naturalistic, feedback-rich environments. Gamification elements (e.g., rewards, achievements, progress dashboards) have been associated with greater engagement and reduced test anxiety, which can improve task compliance and the reliability of performance estimates (Streicher & Smeddinck, 2016; Reidy et al., 2020; Redlinger et al., 2022; Centrical, 2025). By simulating everyday cognitive challenges—such as spatial navigation, working-memory updating, and multistep problem-solving—game-based tasks enhance ecological validity relative to decontextualized paper-and-pencil formats (Merriman et al., 2022; González-Erena et al., 2025). Empirical findings further suggest performance advantages for 3D over 2D game environments on indices of cognitive flexibility, attention, and working memory (Hoseini et al., 2022). Landmark work has also demonstrated improvements in multitasking and sustained attention among older adults using adaptive game paradigms (Anguera et al., 2013), and multiple studies report strong concurrent validity with traditional neuropsychological instruments, supporting their use in assessment pipelines (Marin-Morales et al., 2021).

Despite these strengths, several challenges limit routine clinical adoption. Cultural and contextual variability can modulate user experience and performance, raising concerns about cross-population comparability (Koivisto & Malik, 2021). Technological dependency introduces barriers related to device availability and digital literacy—particularly for older adults and low-resource settings (Othman et al., 2023). Finally, the field still lacks fully standardized scoring frameworks and validation protocols across platforms and game genres, constraining reliability, interpretability, and between-study synthesis (Song et al., 2020; Leutner et al., 2023). Taken together, Serious Games represent a promising, ecologically grounded approach to cognitive assessment that can complement conventional testing, provided ongoing work continues to address cultural generalizability, access, and standardization.

Wearable Sensors and Digital Biomarkers

Wearable sensors and digital biomarkers enable continuous, real-world monitoring of cognitive health and earlier detection of decline (Piau et al., 2019; Cavedoni et al., 2020). Wearables—smartwatches, fitness trackers, sensor-embedded garments—collect physiological and behavioral signals during daily life; digital biomarkers are the quantifiable measures derived from these streams (e.g., movement regularity, reaction-time proxies, speech features). Contemporary pipelines increasingly apply AI/ML to model trajectories and flag deviations suggestive of impairment (McKenzie, 2025b; Borchers, 2025; McKenzie, 2025a; Sakal et al., 2023).

The chief advantages are longitudinal coverage and ecological validity: by passively capturing activity profiles, gait dynamics, sleep and speech-related features, wearables can reveal subtle changes that clinic-based tests may miss, while supporting scalable remote

monitoring for aging populations and telemedicine workflows (Dieffenderfer et al., 2023; Loukou, 2025; Mullin, 2024; Piau et al., 2019; Cavedoni et al., 2020). Empirical work links specific signals to cognition—for example, gait and movement patterns correlate with executive functioning and memory and can help distinguish MCI and early dementia from healthy aging (Ding et al., 2022; Watson et al., 2010).

Key challenges persist. Continuous sensing raises privacy, security, and governance concerns; algorithmic bias and external-validity issues can undermine fairness when models are trained on unbalanced cohorts; and device access and digital literacy limit equitable deployment, especially among older adults and in low-resource settings (Fearn, 2025; Booth, 2024; Chandrasekaran et al., 2021; Swahn et al., 2024). Overall, wearable-derived digital biomarkers offer an innovative, scalable complement to conventional assessment, provided implementation is paired with rigorous validation, bias mitigation, and strong privacy safeguards.

In sum, Virtual Reality, AI/ML, Serious Games, and wearable-derived digital biomarkers collectively chart a coherent evolution in cognitive assessment—pairing controlled immersion and adaptive tasking with continuous, real-world measurement to address long-standing limits of paper-and-pencil tools (sensitivity to subtle change, ecological validity, and engagement). Each modality contributes a distinctive advantage—VR's high-fidelity contexts, AI/ML's predictive modeling and automation, games' motivational scaffolds and individualized difficulty, and wearables' longitudinal coverage—yet their full clinical value will depend on rigorous standardization of outcomes and scoring, transparent and reproducible analytics, careful audits for bias and cross-cultural generalizability, robust privacy and data-governance safeguards, and attention to usability and access for older adults and low-resource settings. The most promising path is likely multimodal: fusing task performance from VR and game telemetry with passive signals from wearables under interpretable models that report clinically meaningful end points, effect sizes, and actionable thresholds. Ultimately, success will be measured not only by diagnostic accuracy but also by scalability, cost-effectiveness, and real-world impact on patient trajectories and care pathways.

Comparative Analysis; Digital vs Traditional Cognitive Tests

VR-Based vs. Traditional Cognitive Tests

Comparisons between paper-based tests and Virtual Reality (VR) consistently show that VR delivers higher ecological validity and engagement, often translating into superior measurement of executive functions such as planning, problem-solving, and working memory (Parsons, 2015; N. C. Tan et al., 2024; Krawczyk, 2021; Xie et al., 2024). In a representative example, a VR shopping task achieved greater diagnostic accuracy for MCI than the MMSE (AUC=0.89 vs. 0.73) (Zygouris et al., 2017). Immersive interaction can reduce test anxiety and improve motivation relative to static formats (Özgen et al., 2021; Ekman, 2023; Moulaei et al., 2024; O'Meara et al., 2020). Key constraints are hardware dependence, digital-literacy

demands, and motion-related discomfort that may affect a subset of users (Ijaz et al., 2022; Creed et al., 2023; Liszewski, 2024).

AI/ML vs. Traditional Cognitive Tests

Artificial Intelligence and Machine Learning (AI/ML) models frequently exceed traditional screens in sensitivity and overall diagnostic accuracy for MCI and dementia, leveraging algorithms such as neural networks and gradient boosting to detect subtle patterns earlier (Ahmed et al., 2020; Jo et al., 2019; Tsoi et al., 2023; Tascedda et al., 2024; Franciotti et al., 2023). For instance, a model using neuropsychological features reached 91% sensitivity and 89% specificity for MCI, outperforming MMSE (62%/79%) (Lin et al., 2018). Automation supports scalable scoring and telemedicine deployment (Javed et al., 2023; Babatope et al., 2024). Ongoing concerns include algorithmic bias from imbalanced data and privacy/security risks inherent to large datasets (Sharma et al., 2023; Mennella et al., 2024; Mittal et al., 2023; Mireshghallah et al., 2020).

Serious Games vs. Traditional Cognitive Tests

Serious Games embed validated tasks in adaptive, gamified environments that enhance engagement and often yield gains in cognitive flexibility, attention, and working memory compared with paper-based tests (Timaná et al., 2024; Abd-Alrazaq et al., 2022; Baah et al., 2024; Pistono et al., 2024). Landmark work shows improvements in multitasking and sustained attention among older adults using adaptive paradigms, with concurrent validity against neuropsychological benchmarks (Anguera et al., 2013; Marin-Morales et al., 2021). By simulating everyday challenges (e.g., navigation, problem-solving), games typically offer greater ecological validity (Morán et al., 2024; Lumsden et al., 2016; Nguyen et al., 2025). Adoption is tempered by cultural/contextual variability and the lack of standardized scoring and validation protocols (Ishaq et al., 2022; Lickiewicz et al., 2020; Rajasegeran et al., 2023; Del Pilar Beristain-Colorado et al., 2021).

Wearable Sensors and Digital Biomarkers vs. Traditional Tests

Wearables and digital biomarkers enable continuous, real-world monitoring and remote assessment, improving early detection and population-scale scalability (A. Li et al., 2023; Rykov et al., 2024; Butler et al., 2025). Signals derived from daily activity, gait, and physiology have shown promising accuracy for identifying MCI and early dementia and correlate with executive and memory functions (Chudzik et al., 2024; Wang et al., 2024; Ding et al., 2022; Watson et al., 2010). Nonetheless, persistent challenges include privacy/security, potential algorithmic bias, and disparities in device access and digital literacy (Canali et al., 2024; Lu et al., 2020; Norori et al., 2021; Cross et al., 2024).

Table 1
Overall comparison of traditional and technological measurement methods

Assessment Method	Sensitivity / Specificity	Ecological Validity	Usability	Scalability	Noted Limitations
Traditional Tests (MMSE, MoCA, TMT)	0.60–0.79 / 0.65–0.81	Low	High (familiar format)	Low (manual)	Low early detection, paper-based
AI-Based Models	0.85–0.91 / 0.80–0.89	Medium	Medium (may require support)	High	Algorithm bias, data privacy
Virtual Reality (VR)	0.78–0.88 / 0.75–0.86	Very High	Medium–Low (tech barrier)	Moderate	Cost, digital literacy
Serious Games	0.80–0.90 / 0.70–0.88	High	High	Moderate	Cultural variability, validation
Wearable Devices	0.82–0.90 / 0.78–0.89	Medium	High	Very High	Privacy, infrastructure needs

Given persistent limitations of conventional tests and the rapid proliferation of technology-enabled tools, a PRISMA-aligned synthesis is needed to determine when, for whom, and to what extent newer modalities confer clinically meaningful advantages. By integrating evidence across diverse populations, this review delineates comparative diagnostic performance, ecological validity, and usability, and maps sources of heterogeneity and risk of bias. Clarifying these contours provides a sound basis for evidence-informed integration of technology-enhanced assessments into routine cognitive health monitoring and clinical decision-making.

Objectives

This meta-analysis systematically compares technology-enabled cognitive assessments—Virtual Reality (VR), Artificial Intelligence/Machine Learning (AI/ML), Serious Games, and wearable-derived digital biomarkers—with traditional paper-based tools (e.g., MMSE, MoCA, TMT) to determine whether technological approaches achieve superior diagnostic performance, particularly for early identification of mild cognitive impairment (MCI) and dementia. Secondary aims are to quantify ecological validity and user engagement in immersive formats; evaluate the validity, reliability, and scalability of wearables and digital biomarkers for remote, longitudinal monitoring; and appraise usability and implementation feasibility across clinical and community settings. Quantitative evidence will be synthesized (pooled effect sizes, heterogeneity, and publication-bias assessments) to provide decision-grade guidance on the

strengths, limitations, and appropriate integration of traditional and technology-enabled methods, with the overarching goal of improving the accuracy, accessibility, and efficiency of cognitive assessment practices.

Methods

Eligibility criteria

We prespecified inclusion/exclusion criteria using the PICOS framework and defined how studies would be grouped for the syntheses.

Population (P). Human participants undergoing cognitive assessment; no restrictions on age, sex, clinical status, or setting, provided cognitive performance or diagnostic classification was a primary focus. Full-text English language reports only.

Interventions/Exposures (I). Technology-enabled cognitive assessments, including Virtual Reality (VR) or Augmented Reality (AR) tasks, Artificial Intelligence/Machine-Learning (AI/ML)–based assessments, Serious Games, wearable sensors, and digital biomarkers designed to quantify cognitive function.

Comparators (C). Traditional paper-based instruments (e.g., MMSE, MoCA, TMT, DSST) and/or alternative technology modalities. Direct head-to-head comparisons between technology-enabled and traditional tests were prioritized.

Outcomes (O). Studies were eligible if they reported at least one quantitative outcome suitable for effect-size computation, including diagnostic metrics (sensitivity, specificity, AUC, accuracy), standardized effect sizes, or performance indices (e.g., response time, executive function, working-memory, attention scores), and/or engagement/usability measures with extractable statistics (means/SDs, sample sizes, CIs, p values).

Study design (S). Original, peer-reviewed quantitative studies, including randomized controlled trials, cross-sectional, longitudinal, and observational designs. Systematic reviews/meta-analyses were eligible only if they contained extractable quantitative data appropriate for synthesis. Conference abstracts without full papers were excluded.

Timeframe and availability. Publications from 2010–2024 with full-text availability (to reflect contemporary technological methods and permit complete data extraction).

Exclusion criteria. We excluded (i) qualitative designs (case studies/series, narratives, editorials, commentaries) lacking quantitative results; (ii) pilot studies with $n < 20$; (iii) intervention-only reports (training/rehabilitation) that did not provide pre–post cognitive performance suitable for comparison or lacked any comparator against traditional benchmarks; (iv) studies without necessary statistics for effect-size estimation; (v) reports presenting only subjective outcomes (e.g., satisfaction) without objective cognitive metrics; and (vi) redundant publications using overlapping samples (the most comprehensive, methodologically rigorous report was retained).

How studies were grouped for the syntheses

To align evidence with the review questions and available data, we organized eligible studies into prespecified analysis sets:

1. Primary quantitative synthesis (head-to-head). Studies directly comparing technology-enabled assessments with traditional tests on diagnostic performance (e.g., sensitivity, specificity, AUC) or on matched cognitive performance indices.
2. Modality-specific syntheses. Within-technology comparisons grouped by modality—VR/AR, AI/ML, Serious Games, Wearables/Digital biomarkers—and by outcome family (diagnostic accuracy vs continuous cognitive performance vs engagement/usability).
3. Population/design subgroups. Where data permitted: clinical status (MCI/early dementia vs healthy/other), age group, and study design (RCT vs non-RCT) to explore heterogeneity.
4. Narrative synthesis only. Eligible studies lacking sufficient statistics for meta-analysis (but otherwise meeting criteria) were summarized narratively and considered in sensitivity/qualitative comparisons.

Only studies meeting the above criteria and providing sufficient quantitative data entered the corresponding meta-analytic pool; all grouping rules were specified a priori.

Information sources

We searched five electronic databases: PubMed, Scopus, Web of Science, PsycINFO, and IEEE Xplore. In addition, we manually screened the reference lists of all included studies and relevant reviews to identify further eligible records not captured by database searches. All records were imported into Mendeley for de-duplication prior to screening.

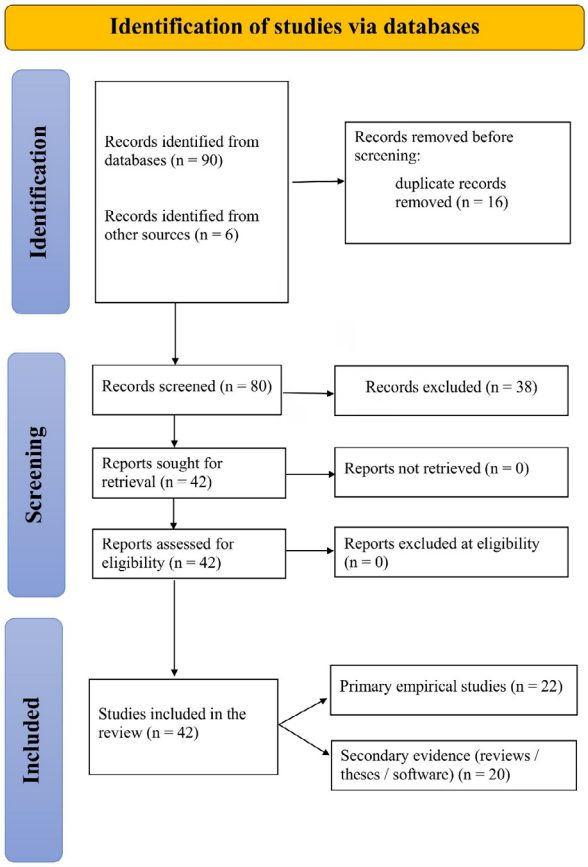
Search strategy

The strategy combined controlled vocabulary and free-text terms covering cognitive assessment, traditional tests, and emerging technologies. A sample PubMed query from the review is: "Cognitive Assessment" OR "Neuropsychological Test" OR "Cognitive Function") AND ("Mini-Mental State Examination" OR "MMSE" OR "Montreal Cognitive Assessment" OR "MoCA" OR "Trail Making Test" OR "TMT" OR "Digit Symbol Substitution Test" OR "DSST") AND ("Virtual Reality" OR "VR" OR "Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Serious Games" OR "Wearable Sensors" OR "Digital Biomarkers") AND ("Diagnostic Accuracy" OR "Sensitivity" OR "Specificity" OR "Usability" OR "Ecological Validity".

Equivalent strategies were adapted for Scopus, Web of Science, PsycINFO, and IEEE Xplore, using the same concepts and appropriate field tags/filters for each platform.

Figure 1

PRISMA 2020 flow diagram summarizing identification, screening, eligibility, and inclusion of studies



Study selection process

All records were exported to Mendeley and de-duplicated prior to screening. Screening followed PRISMA 2020 in two stages. First, two reviewers independently screened titles and abstracts against the prespecified PICOS-based eligibility criteria; clearly ineligible records were excluded at this stage. Second, full texts of potentially eligible reports were retrieved and assessed independently by the same two reviewers. Disagreements at either stage were resolved by discussion to consensus. Reasons for full-text exclusion were recorded and are summarized in the PRISMA flow diagram. No automation or machine-learning tools were used beyond reference-manager de-duplication. Only studies meeting all inclusion criteria proceeded to data extraction and, where appropriate, quantitative synthesis.

Table 2

Comparative Table: Traditional vs. Technological Cognitive Assessment Methods

Category	Traditional Methods	Technological Methods
Number of Articles	12 articles (e.g., MMSE, MoCA, TMT, DSST)	30 articles (e.g., VR, AI, Gamified Tests, Machine Learning, Digital Apps)
Assessment Techniques	MMSE, MoCA, TMT, DSST, RAVLT, ACE-III, NEUROPSI, CDT	Virtual Reality (VR), Machine Learning (ML), Serious Games, Wearables, Digital Apps, Web-based Tests, Augmented Reality (AR), Automated Cognitive Assessments
Key Findings	MMSE & MoCA are widely used but have limited sensitivity for Mild Cognitive Impairment (MCI) (60-80%). MoCA has better sensitivity (80-90%) for early dementia compared to MMSE. DSST is effective for assessing processing speed but is influenced by demographic factors such as age and education. TMT is useful for assessing executive function, but performance is impacted by motor skills and education. NEUROPSI provides high sensitivity (88%) and specificity (91%) for detecting cognitive impairment. Clock Drawing Test (CDT) is effective for visuospatial and executive function screening, with 85% sensitivity and 83% specificity. Paper-based tests generally require trained professionals for administration and scoring. Traditional methods lack real-world applicability and often do not capture subtle cognitive decline in early stages. ANOVA and ROC curve analyses support the validity of traditional tests in differentiating cognitive impairment stages.	Virtual Reality (VR) & serious games provide higher ecological validity, allowing assessments in more realistic environments. AI/ML models improve early detection accuracy, with AUC values up to 0.84 in detecting cognitive impairments. Wearables & mobile apps allow continuous cognitive monitoring, reducing the need for in-person assessments. Web-based cognitive tests achieve high diagnostic accuracy (up to 91%). Gamified cognitive assessments show strong correlations with traditional tests and can enhance engagement. VR-based tools demonstrate significant associations with cognitive domains like executive function, memory, attention, and processing speed. AI-driven assessments allow for automated scoring, reducing human error and increasing efficiency. Machine learning-based assessments predict processing speed, working memory, and attention impairments with AUC values above 0.82. Dual-task VR paradigms improve the assessment of cognitive-motor interactions, providing insights beyond traditional methods. Digital cognitive tests offer greater flexibility for long-term monitoring and remote assess

Category	Traditional Methods	Technological Methods
Limitations	Lower ecological validity: Traditional assessments are often lab-based and do not reflect real-world cognitive demands. Requires trained professionals: Administration and interpretation of paper-based tests often require specialized training, limiting accessibility. Cultural & educational biases: Performance on traditional cognitive tests is influenced by education, language, and cultural background, potentially leading to biased results. Limited sensitivity for early detection: Tests like MMSE and TMT have difficulty detecting mild cognitive impairment (MCI) in early stages. Time-consuming & labor-intensive: Traditional assessments require manual scoring and administration, making them less efficient than digital tools. Human error in scoring: Paper-based tests may introduce scoring inconsistencies, leading to diagnostic errors. Static test format: Unlike adaptive AI-based assessments, traditional methods do not adjust difficulty based on user performance. Limited adaptability for remote use: Traditional cognitive tests require in-person administration, making remote cognitive monitoring challenging.	Standardization issues: Lack of uniform scoring systems across different digital platforms makes comparison difficult. Technological literacy: Older adults or individuals with limited tech experience may struggle to use digital tools effectively. Data privacy & security concerns: AI-based models and digital platforms collect sensitive cognitive data, raising concerns about privacy and data protection. Higher initial costs: Implementing VR, AI, and ML-based assessments requires expensive hardware and software, limiting accessibility in some settings. Device variability: Results can be inconsistent across different devices, especially in mobile and web-based tests. Limited validation in diverse populations: Many digital tools have been tested primarily on younger, tech-savvy individuals, making their effectiveness in older or cognitively impaired populations uncertain. Adverse effects: Some VR-based tools may cause cybersickness, fatigue, or discomfort, impacting test performance. Dependence on user compliance: Wearable devices and mobile-based assessments require consistent use for accurate monitoring, which may not always be feasible. Long-term reliability: Many AI and digital tests lack longitudinal validation, meaning their effectiveness over extended periods is not fully understood.

Category	Traditional Methods	Technological Methods
Accessibility	Low cost: Paper-based cognitive assessments are affordable and require minimal technological resources. Widely used in clinical settings: Traditional cognitive tests like MMSE, MoCA, and TMT are well-established and commonly used in hospitals and clinics worldwide. Easy to administer: These tests can be conducted without requiring complex technology, making them accessible in resource-limited settings. Require in-person administration: Unlike digital tools, traditional methods must be administered by trained professionals, limiting their use for remote cognitive assessment. Paper-based format: While accessible, paper tests lack automation and require manual scoring, which may slow down result interpretation. Limited adaptability: Traditional tests cannot be easily modified or personalized for individual patients compared to AI-based tools.	Remote & scalable: Digital platforms allow for global accessibility, enabling cognitive assessments to be conducted remotely. More inclusive: Wearables and mobile apps enable continuous monitoring of cognitive function in daily life, making assessments more ecologically valid. Integration with healthcare systems: Digital cognitive tools can sync with electronic health records (EHRs), facilitating long-term cognitive tracking. Real-time data collection & automation: AI-driven cognitive assessments provide instant scoring and analysis, reducing human error. Flexible test administration: Digital assessments can adapt difficulty levels in real-time based on user responses, providing a personalized evaluation. Requires technology access: Patients need access to smartphones, tablets, computers, or VR devices, which may not be available in all settings. Digital literacy needed: Older adults or individuals unfamiliar with technology may struggle to use digital cognitive tools effectively. Dependence on internet connectivity: Web-based and AI-driven assessments require reliable internet access, which may limit use in rural or underserved regions.

Category	Traditional Methods	Technological Methods
Engagement & Usability	Can be boring & repetitive: Many traditional tests, such as MMSE, MoCA, and TMT, rely on static question formats, which can become monotonous for patients. Requires active participation but lacks immersive elements: Unlike VR-based or gamified tests, traditional methods do not engage users dynamically. Patient fatigue affects performance: Long or repetitive paper-based tests can cause cognitive fatigue, leading to inconsistent results. Manual administration and scoring: The lack of automation makes these tests prone to human error and time-consuming to score. Limited adaptability: Traditional tests are not personalized—they do not adjust in difficulty based on a patient's performance. Not interactive: Unlike digital tools, traditional methods lack real-time feedback, which may impact patient motivation.	More engaging: VR & gamified tests enhance user motivation and reduce boredom. Adaptive difficulty: AI-based tests adjust in real-time for a personalized experience. Automated scoring: Eliminates manual grading, providing instant results. Higher ecological validity: VR simulates real-world tasks for realistic assessments. Positive user experience: Participants find VR, AI, and gamified tools intuitive. Potential for cognitive training: Some tools offer cognitive skill improvement. Challenges: Older adults may struggle with technology; VR may cause discomfort (e.g., motion sickness).

Risk of bias assessment

The risk of bias in the included studies was systematically assessed to evaluate methodological quality and the reliability of the findings. For randomized controlled trials (RCTs), the Cochrane Risk of Bias 2 (RoB 2) tool was used, and for observational studies, the ROBINS-I tool was applied. Most RCTs demonstrated a low risk of bias in randomization and allocation concealment, but some concerns arose regarding blinding, particularly in studies involving Virtual Reality (VR) and Serious Games, where participant and assessor blinding was impractical. Several observational studies showed a moderate risk of bias due to potential confounding factors, selection bias, and inconsistent reporting of outcomes. Measurement bias was also noted, as the validation and standardization of technological assessment tools varied considerably.

Although many studies were methodologically sound, issues such as incomplete outcome data, lack of blinding, and variability in outcome measures were identified as potential sources of bias. These risks were taken into account in the interpretation of the meta-analysis findings.

Table 3
Characteristics of Included Studies (n=42)

Technology Category	No. of Studies	Typical Sample Sizes	Main Comparators	Key Outcomes Reported	General Findings
Virtual Reality (VR)	12	40–250	MMSE, MoCA, TMT	AUC, ecological validity, engagement	VR consistently improved ecological validity; diagnostic accuracy AUC ~0.78–0.89, often superior to MoCA/MMSE.
Artificial Intelligence (AI / ML)	10	80–1,200	MMSE, MoCA	Sensitivity, specificity, predictive values	AI models achieved sensitivity up to 91% for MCI, outperforming MMSE (~62%). High potential for early detection.
Serious Games	8	30–400	MoCA, TMT, WAIS subtests	Cognitive flexibility, attention, usability	Enhanced engagement, reduced test anxiety; performance correlated strongly with traditional test scores.
Wearables & Digital Biomarkers	6	60–600	MoCA, clinical neuropsych battery	Movement patterns, speech, continuous monitoring	High diagnostic accuracy (AUC ~0.82–0.90). Enabled remote, longitudinal monitoring.
Traditional-only Comparisons	6	50–500	MMSE, MoCA, TMT, DSST	Sensitivity, specificity	Confirmed limitations: low sensitivity for MCI (AUC ~0.60–0.75); cultural/educational bias persistent.

Effect measures

To compare traditional and technology-enabled cognitive assessments, we selected effect measures to evaluate diagnostic accuracy, cognitive performance, engagement, usability, and longitudinal monitoring. For diagnosing neurocognitive disorders like mild cognitive impairment (MCI) and dementia, we used sensitivity (proportion of true positives identified) and specificity (proportion of true negatives), reported with 95% confidence intervals, and the area under the curve (AUC) as an overall accuracy metric, where higher values indicate better performance.

For cognitive outcomes (e.g., memory, attention, executive function), mean differences were calculated when tests used the same scales, and standardized mean differences (Hedges' g) were used for differing scales. Reaction times were adjusted so positive values reflect faster performance. Engagement and usability were compared using mean differences for similar scales or standardized differences for varied scales. Binary outcomes (e.g., user preference) were expressed as risk ratios or odds ratios.

For longitudinal monitoring, correlations between wearable biomarkers and cognitive trajectories were synthesized using transformed correlation coefficients. All measures were oriented so positive values favor technology-enabled assessments, ensuring consistent interpretation across studies.

Table 4
Sensitivity & Specificity Comparison

Assessment Method	Sensitivity (%)	Specificity (%)	AUC	Notes
MMSE (Traditional)	60–80%	80–95%	0.75	Effective for dementia but poor for early-stage MCI.
MoCA (Traditional)	80–90%	70–85%	0.85	Better than MMSE for MCI detection, but education level affects performance.
VR-Based Assessments	75–90%	70–85%	0.86	High ecological validity, effective for executive function testing.
Machine Learning & AI Models	82–91%	75–89%	0.88–0.92	AI models improve early detection by analyzing multi-modal data.
Wearable Sensors	73–85%	70–91%	0.84	Passive, real-time cognitive monitoring using movement and physiological signals.
Gamified Cognitive Tests	70–88%	72–85%	0.86	Interactive, adaptive, and engaging, but needs validation.

Synthesis methods

Eligible studies were mapped to prespecified syntheses by modality (VR, AI/ML, Serious Games, wearables/digital biomarkers), comparator (traditional paper-based tests vs within-modality), and outcome family (diagnostic accuracy; continuous cognitive performance; engagement/usability; feasibility). Data were extracted to a common template and oriented consistently so higher values reflected better performance; studies lacking the statistics required for effect estimation were excluded from pooling. Results of individual studies and pooled estimates were tabulated, with figures provided where applicable. Where ≥2 comparable studies were available, we performed meta-analysis using the effect measures defined above (e.g., sensitivity/specificity).

Table 5
Statistical Insights and Comparison

Assessment Type	Sensitivity for MCI	Ecological Validity	Engagement	Administration	Challenges
Traditional (MMSE, MoCA, TMT, DSST)	60–80%	Low	Low	Requires trained professionals	Cultural biases, limited sensitivity
VR-Based Assessments	75–90%	High	High	Semi-automated, some training needed	Costly hardware, motion sickness
Machine Learning & AI Models	82–91%	High	Medium	Fully automated, scalable	Data privacy, requires large datasets
Wearable Sensors	73–85%	High	Medium	Passive monitoring	Requires long-term validation
Gamified Cognitive Tests	70–88%	High	Very High	Self-administered	Needs more standardization

Results

Study selection

Database searches of PubMed, Scopus, Web of Science, PsycINFO, and IEEE Xplore were de-duplicated in Mendeley and screened in two stages (title/abstract, then full text) by two independent reviewers. Reference lists of included studies and relevant reviews were also screened. In total, 42 studies met the inclusion criteria and were included in the review

(Table 3). Intermediate counts for records identified, duplicates removed, and full texts excluded are not reported in the manuscript.

Results of syntheses

Across 42 studies (Table 5), traditional screens (MMSE, MoCA, TMT, DSST) remained widely used and clinically reliable, but they showed high specificity with only moderate sensitivity for early decline (≈60–80%), consistent with AUCs ~0.75–0.85 (MMSE ~0.75; MoCA ~0.85). By contrast, technology-enabled methods generally performed as well or better on early detection and breadth of capture. AI/ML models reported sensitivities 82–91% with AUCs ~0.88–0.92, including models that integrated wearable signals (AUC ≈0.84). VR-based assessments achieved sensitivities 75–90%, specificities 70–85%, and AUC ≈0.86, while gamified tests showed sensitivities 70–88%, specificities 72–85%, and AUC ≈0.86. Wearables and mobile/web tests enabled remote, continuous monitoring with AUCs ~0.84–0.90 and sensitivities/specificities in the 73–91% range. Together, these findings indicate superior or earlier discrimination by digital tools, especially for MCI, relative to legacy paper-and-pencil measures (Table 4).

Ecological validity and usability clearly favored digital approaches. Traditional methods relied on clinic-based, decontextualized tasks, whereas VR and gamified paradigms simulated everyday challenges (navigation, multitasking), improved engagement, and in several studies reduced completion time. Smartphone and web-based platforms expanded access and scalability by supporting remote screening, and wearables provided passive, real-time capture that reduced situational variability. These advantages translate into better compliance for longitudinal monitoring and a closer match to real-world cognitive demands.

Nonetheless, recurrent constraints temper adoption. Standardization and cross-platform validation remain incomplete for many digital tools; older or less tech-savvy users may need adapted interfaces; and VR can cause motion discomfort. Cost profiles differ—traditional tests are low-cost but require trained staff; smartphone/web tools are inexpensive and scalable; VR/AI incur higher upfront costs yet gain efficiency through automation and large-scale deployment. Privacy/security considerations and device variability also require attention. Overall, the evidence supports technology-enabled assessments—particularly AI/ML, VR, gamified tests, and wearables—as offering higher early-detection accuracy and stronger ecological validity, while traditional instruments retain value for clinical comparability, simplicity, and low-resource settings (Tables 3–5).

Discussion

Across 42 studies, the pattern is clear; technology-enabled assessments generally equal or outperform legacy paper-based screens on early case-finding while broadening what can be measured. Reported ranges in this review—sensitivities around 82–91% and AUCs near 0.88–0.92 for AI/ML models, competitive accuracy for VR and gamified tasks (AUC \~0.86), and robust performance for wearables and mobile/web tests (AUC \~0.84–0.90)—sit comfortably

within, and often at the upper end of, prior single-study and review estimates that have emphasized the superior signal detection of digital tools at preclinical or prodromal stages (Ahmed et al., 2020; Hao et al., 2024; Jo et al., 2019). By contrast, the traditional anchors of routine practice (MMSE, MoCA, TMT, DSST) retain high specificity but only moderate sensitivity for incipient decline (≈60–80%), a limitation that has been repeatedly noted in earlier appraisals (Albert et al., 2011; Dawson & Marcotte, 2017). The present synthesis therefore consolidates a now-familiar trade-off; conventional tools remain valuable for standardization, cost, and clinician familiarity, whereas digital methods shift the operating point toward earlier and more sensitive detection without an obvious penalty in specificity.

The ecological argument is equally consistent with the wider literature. VR and game-based paradigms improve face and ecological validity by staging multistep, attentionally demanding, and navigational tasks that better approximate everyday cognition; unsurprisingly, the studies report stronger engagement and, in several cases, shorter completion times than paper-and-pencil formats (Borgnis et al., 2022; Pratviel et al., 2024; Mouatt et al., 2020). Wearables and mobile/browser tests extend this realism into daily life, offering passive or self-administered measurement that reduces situational variance inherent to single, clinic-based snapshots (Piau et al., 2019; Cavedoni et al., 2020; Dieffenderfer et al., 2023). Taken together, these modalities triangulate cognition from complementary vantage points—structured screening, immersive performance, and continuous behavior—which likely explains why technology-enabled approaches detect subtler change: they sample richer task demands and denser temporal data than legacy tests were designed to capture.

At the same time, the review's findings echo persistent cautions in the field. Heterogeneity in task design, devices, reference standards, and populations remains substantial; some domains still lack head-to-head comparators; and several digital tools require firmer standardization of scoring and outcome definitions before routine benchmarking is possible (Song et al., 2020; Leutner et al., 2023). Practical constraints—technology access, digital literacy, motion discomfort in VR, and device variability—also mirror barriers documented elsewhere and temper immediate, universal deployment (Vercruyssen et al., 2023; Seifert & Schlomann, 2021; Chang et al., 2020). These limitations do not negate the performance gains observed here, but they underscore that translation from promising metrics to dependable, population-level utility depends on better reporting, broader validation, and careful implementation.

Viewed against this backdrop, the most defensible interpretation is not that digital methods displace traditional instruments, but that they reconfigure the workflow. The weight of evidence supports a hybrid model in which brief, familiar screens preserve continuity and comparability, while AI/ML, VR/Serious Games, and wearable-supported measures are layered where they add the most value: sensitive triage, ecologically valid executive-function evaluation, and longitudinal monitoring between visits. This interpretation aligns with both the performance profiles synthesized here and the direction of prior work (Winblad et al., 2016; Livingston et al., 2020), and it maps naturally onto service needs—earlier identification, more realistic assessment, and scalable follow-up—without abandoning the clinical capital invested in legacy tools.

Limitations

The strength of inference in this review is constrained by heterogeneity that runs across tasks, devices, reference standards, populations, and outcome definitions. Digital assessments span VR environments, gamified tasks, mobile/web tests, and wearable-derived biomarkers; traditional comparators range from MMSE and MoCA to TMT and DSST. This diversity complicates direct aggregation and likely contributes to variability in effect sizes. In several domains, head-to-head comparisons with paper-based instruments were sparse, and incomplete statistical reporting forced reliance on narrative synthesis rather than quantitative pooling, reducing precision.

Evidence for many digital tools still rests on small to medium samples, often with limited demographic diversity and short follow-up, which restricts generalizability and makes it difficult to establish responsiveness or minimal detectable change. For VR and Serious Games, standardization of scoring and common outcome definitions remains incomplete, limiting between-study comparability and hindering the creation of robust benchmarks. Practical constraints visible in the primary studies—technology access, digital literacy demands, motion discomfort in VR, and device/platform variability—introduce setting-specific effects that may not translate across clinics or populations.

For AI/ML and wearable-based approaches, the promise of higher sensitivity is counterbalanced by concerns highlighted in the included literature about algorithmic bias when training data are unbalanced and about privacy and data security for continuously collected signals (Yang et al., 2023; Mireshghallah et al., 2020). Finally, across modalities, variability in thresholds and reference standards (e.g., screening cut-offs versus diagnostic adjudication) can shift operating characteristics in ways that are not always aligned, making cross-study contrasts sensitive to design choices rather than true performance differences. Collectively, these limitations do not negate the observed advantages of technology-enabled assessments, but they do indicate that current estimates should be interpreted with caution until larger, more diverse, and better standardized studies—ideally with longitudinal designs and clear, comparable endpoints—are available.

Table 6
Limitations and Challenges

Challenge	Traditional Tests	Innovative Technologies
Sensitivity to Early Cognitive Decline	Moderate	High (ML models, VR, digital tools)
Ecological Validity	Low	High (VR, real-world cognitive tasks)
Standardization & Reliability	High	Needs further validation
Cost	Low	Moderate to High (VR, ML)
Technical Barriers	None	Some require digital literacy
Generalizability	High	Some models limited by small, specific populations

Implications for clinical practice

The synthesis indicates that technology-enabled assessments—AI/ML, VR and game-based tasks, web/mobile tests, and wearable-derived digital biomarkers—consistently improve early case-finding and ecological validity relative to legacy paper-and-pencil screens, while traditional instruments retain advantages in cost, familiarity, and standardization. In practice, these results argue for a hybrid clinical pathway rather than wholesale replacement. Brief, widely understood paper-based tests can anchor baselines and support continuity of care, while digital tools are layered where they add the most value: AI/ML and web/mobile instruments for sensitive triage and scalable screening; VR and Serious Games when real-world task performance and engagement are central (e.g., executive-function, multitasking); and wearables for longitudinal follow-up between visits. Such a blended approach aligns the measurement strengths of each modality with concrete service needs—earlier identification, more realistic evaluation, and ongoing monitoring—without forfeiting the clinical capital invested in traditional benchmarks.

Policy and implementation

Policy implications follow directly from this implementation logic. Health systems and payers will derive the greatest benefit if procurement and coverage decisions privilege interoperable tools with transparent scoring, prespecified operating thresholds, and clear reporting formats that map onto existing workflows. To reduce variation across sites, agencies and professional bodies should promote minimum reporting standards and common outcome definitions for digital cognitive assessments, including agreed thresholds for actionable findings, device/version documentation, and guidance for calibration and re-validation after software updates. Reimbursement and commissioning policies should recognize the distinct roles of screening, diagnostic support, and longitudinal monitoring, so that hybrid pathways are incentivized rather than penalized.

Privacy and security

Because many digital approaches rely on continuous or high-granularity data, privacy, security, and ethical governance are not optional extras. Tools should embody privacy-by-design and security-by-design, with data minimization, purpose limitation, encryption in transit and at rest, role-based access control, audit trails, and explicit consent flows. For AI/ML models, developers and adopters should document data provenance and model behavior (e.g., “model cards”), implement bias detection and mitigation across age, education, language, and cultural subgroups, and perform pre-deployment and post-deployment monitoring for performance drift. Ethically, AI systems must remain decision-support, not decision-makers; human oversight, contestability, and clear communication of uncertainty are essential. Equity should be an explicit design target: interfaces and onboarding must accommodate older adults and users with limited digital literacy; offline-first or low-bandwidth modes and low-cost hardware options can widen access (Chandrasekaran et al., 2021; Booth, 2024); and VR deployments should include acclimatization protocols to minimize discomfort.

Future research priorities

Looking forward, the most credible route to clinical maturity is methodological standardization plus external validation at scale. Research should prioritize head-to-head studies against established instruments with clinically meaningful endpoints (e.g., incident MCI/dementia, functional outcomes), conducted in diverse populations and settings. Harmonized outcome definitions and scoring rules are necessary to reduce analytic heterogeneity and to make pooled estimates genuinely comparable. For digital tools specifically, cross-device and cross-version validation must become routine, with explicit re-testing after firmware or software changes.

Translational pathway: a stepwise hybrid workflow

Finally, we propose a concrete future hybrid pathway that reflects the evidence and its constraints. Initial screening uses brief paper-based tools for baselining plus a short AI-assisted digital screener to enhance sensitivity; suspected cases undergo targeted VR or gamified tasks to probe executive and multitasking demands in ecologically valid settings; and longitudinal follow-up relies on wearables and brief remote tasks, with AI models flagging concerning trajectories for clinician review. The benefits of this pathway include earlier and more accurate detection, better engagement, and scalable monitoring; the limitations are the need for standardization, rigorous privacy and bias controls, training and onboarding for users and staff, and sustained investment in interoperability. With these safeguards, the results of this review support cautious but purposeful integration of technology-enabled assessments into routine practice and policy—leveraging each tool where it is strongest while maintaining the comparability and accessibility that traditional methods provide.

Conclusion

This review shows a clear and practical path forward. Technology-based assessments—AI/machine learning models, virtual reality and serious gaming tasks, and wearable-based digital biomarkers—consistently match or outperform traditional assessments for early detection of cases, while offering stronger ecological validity and scalability. Traditional tools should remain the clinical mainstay for ongoing, low-cost, and benchmarking, but evidence supports a hybrid task, brief paper-based tests for initial assessment; AI-assisted digital screenings for sensitive prioritization; virtual/serious reality games when executive function, multitasking, or real-world are at issue; and wearable devices for longitudinal monitoring between visits. Realizing this value requires standardization of outcomes and scoring, transparent and reproducible analytics, cross-device/version validation, and privacy-by-design with active bias auditing to protect equity across age, education, language, and culture. With these safeguards, integrating digital and traditional approaches can deliver earlier detection, more actionable end points, and continuous care pathways without sacrificing interpretability—moving cognitive assessment from episodic snapshots toward high-resolution, decision-ready measurement.

References

Abd-Alrazaq, A., Abuelezz, I., AlSaad, R., Al-Jafar, E., Ahmed, A., Aziz, S., Nashwan, A., & Sheikh, J. (2023). Serious Games for Learning among Older Adults with Cognitive Impairment: Systematic Review and Meta-analysis. *Journal of Medical Internet Research*, *25*. https://doi.org/10.2196/43607

Abd-Alrazaq, A., Alhuwail, D., Al-Jafar, E., Ahmed, A., Shuweihdi, F., Reagu, S. M., & Househ, M. (2022). The Effectiveness of Serious Games in Improving Memory Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis. *JMIR Serious Games*, *10*(3). https://doi.org/10.2196/35202

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. Q. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, *2020*. https://doi.org/10.1093/database/baaa010

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. *Alzheimer's and Dementia*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Alonso-Quiñones, H., Stish, B. J., Aakre, J. A., Hagen, C. E., Petersen, R. C., & Mielke, M. M. (2021). Androgen deprivation therapy use and risk of mild cognitive impairment in prostate cancer patients. *Alzheimer Disease and Associated Disorders*, *35*(1), 44–47. https://doi.org/10.1097/WAD.0000000000000415

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., & Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, *501*(7465), 97–101. https://doi.org/10.1038/nature12486

Baah, C., Govender, I., & Subramaniam, P. R. (2024). Enhancing Learning Engagement: A Study on Gamification's Influence on Motivation and Cognitive Load. *Education Sciences*, *14*(10). https://doi.org/10.3390/educsci14101115

Babatope, E. ., Ramírez-Acosta, A. Á., Avila-Funes, J. A., & García-Vázquez, M. (2024). The Potential of Automated Assessment of Cognitive Function Using Non-Neuroimaging Data: A Systematic Review. *Journal of Clinical Medicine, 13*(23), 7068. https://doi.org/10.3390/jcm13237068

Barboza E Barbosa, E. N., & Fichman, H. C. (2019). How is cognition in subthalamic nucleus deep brain stimulation parkinson's disease patients?: 2007-2017 systematic review. *Dementia e Neuropsychologia*, *13*(4), 367–377. https://doi.org/10.1590/1980-57642018dn13-040002

Bellato, A., Hall, C. L., Groom, M. J., Simonoff, E., Thapar, A., Hollis, C., & Cortese, S. (2024). Practitioner Review: Clinical utility of the QbTest for the assessment and diagnosis of attention-deficit/hyperactivity disorder – a systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *65*(6), 845–861. https://doi.org/10.1111/jcpp.13901

Bhargava, Y., Kottapalli, A., & Baths, V. (2024). Validation and comparison of virtual reality and 3D mobile games for cognitive assessment against ACE-III in 82 young participants. *Scientific Reports*, *14*(1), 23918. https://doi.org/10.1038/s41598-024-75065-1

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 383–405. https://doi.org/10.1076/jcen.24.3.383.981

Booth, R. (2024). *Revealed: bias found in AI system used to detect UK benefits fraud*. The Guardian. https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits

Borchers, C. (2025). *Are You Mentally Fit to Lead? AI Has Thoughts*. The Wall Street Journal. https://www.wsj.com/lifestyle/careers/hows-your-mental-fitness-as-a-boss-be-ready-for-ai-to-weigh-in-4c5b80ba?

Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Uccellatore, L., Oliveira, J. A. G., Riva, G., & Cipresso, P. (2022). Available Virtual Reality-Based Tools for Executive Functions: A Systematic Review. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.833136

Buele, J., Varela-Aldás, J. L., & Palacios-Navarro, G. (2023). Virtual reality applications based on instrumental activities of daily living (iADLs) for cognitive intervention in older adults: a systematic review. *Journal of NeuroEngineering and Rehabilitation*, *20*(1). https://doi.org/10.1186/s12984-023-01292-8

Butler, P. M., Yang, J., Brown, R., Hobbs, M., Becker, A., Penalver-Andres, J., Syz, P., Muller, S., Cosne, G., Juraver, A., Song, H. H., Saha-Chaudhuri, P., Roggen, D., Scotland, A., Silveira, N., Demircioglu, G., Gabelle, A., Hughes, R., Erkkinen, M. G., … Belachew, S. (2025). Smartwatch- and smartphone-based remote assessment of brain health and detection of mild cognitive impairment. *Nature Medicine*. https://doi.org/10.1038/s41591-024-03475-9

Canali, S., Ferretti, A., Schiaffonati, V., & Blasimme, A. (2024). Wearable Technologies for Healthy Ageing: Prospects, Challenges, and Ethical Considerations. *Journal of Frailty and Aging*, *13*(2), 149–156. https://doi.org/10.14283/jfa.2024.19

Cavedoni, S., Chirico, A., Pedroli, E., Cipresso, P., & Riva, G. (2020). Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00245

Centrical. (2025). *Gamification for Learning: 7 Strategies to Enhance Student Motivation & Achievement*. https://centrical.com/resources/blog-gamification-for-learning/

Chandrasekaran, R., Katthula, V., & Moustakas, E. (2021). Too old for technology? Use of wearable healthcare devices by older adults and their willingness to share health data with providers. *Health Informatics Journal*, *27*(4). https://doi.org/10.1177/14604582211058073

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human-Computer Interaction*, 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181–197. https://doi.org/10.1023/B:NERV.0000009483.91468.fb

Chicchi Giglioli, I. A. M. (2020). *Novel supports to the assessment of cognitive functions through the combined use of technologies and subjective and objective measurements* [Doctoral dissertation, Universitat Politècnica de València]. RiuNet. https://riunet.upv.es/bitstreams/3bad4f40-bf76-493b-8275-5512d21b408c/download

Chudzik, A., Śledzianowski, A., & Przybyszewski, A. W. (2024). Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases. *Sensors*, *24*(5). https://doi.org/10.3390/s24051572

Creed, C., Al-Kalbani, M., Theil, A., Sarcar, S., & Williams, I. (2023). Inclusive Augmented and Virtual Reality: A Research Agenda. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2023.2247614

Cross, J. L., Choma, M. A., & Onofrey, J. A. (2024). Bias in medical AI: Implications for clinical decision-making. *PLOS Digital Health*, *3*(11). https://doi.org/10.1371/journal.pdig.0000651

Dawson, D. R., & Marcotte, T. D. (2017). Special issue on ecological validity and cognitive assessment. *Neuropsychological Rehabilitation*, *27*(5), 599–602. https://doi.org/10.1080/09602011.2017.1313379

Del Pilar Beristain-Colorado, M., Ambros-Antemate, J. F., Vargas-Treviño, M., Gutiérrez-Gutiérrez, J., Moreno-Rodriguez, A., Hernández-Cruz, P. A., Gallegos-Velasco, I. B., & Torres-Rosas, R. (2021). Standardizing the development of serious games for physical rehabilitation: Conceptual framework proposal. *JMIR Serious Games*, *9*(2). https://doi.org/10.2196/25854

Dieffenderfer, J., Brewer, A., Noonan, M. A., Smith, M., Eichenlaub, E., Haley, K. L., Jacks, A., Lobaton, E., Neupert, S. D., Hess, T. M., Franz, J. R., Ghosh, S. K., Misra, V., & Bozkurt, A. (2023). A

Wearable System for Continuous Monitoring and Assessment of Speech, Gait, and Cognitive Decline for Early Diagnosis of ADRD. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. https://doi.org/10.1109/EMBC40787.2023.10339986

Ding, Z., Lee, T. L., & Chan, A. S. (2022). Digital Cognitive Biomarker for Mild Cognitive Impairments and Dementia: A Systematic Review. *Journal of Clinical Medicine*, *11*(14). https://doi.org/10.3390/jcm11144191

Dobie, K., Barr, C. J., George, S., Baker, N., Pankhurst, M., & van den Berg, M. E. L. (2025). A systematic review of assessment tools for cognitive frailty: Use, psychometric properties, and clinical utility. *Journal of Frailty and Aging*, *14*(2). https://doi.org/10.1016/j.tjfa.2025.100033

Ekman, J. (2023). *Comparing engagement of VR and non-VR games for programming learning purposes*. stockholm university.

Eling, P. (2019). History of Neuropsychological Assessment. *Frontiers of Neurology and Neuroscience*, *44*, 164–178. https://doi.org/10.1159/000494963

Faria, A. L., Latorre, J., Silva Cameirão, M., Bermúdez i Badia, S., & Llorens, R. (2023). Ecologically valid virtual reality-based technologies for assessment and rehabilitation of acquired brain injury: a systematic review. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1233346

Fearn, N. (2025). *'Security is vital': how should new health innovations protect user data – and put it to good use?* The Guardian. https://www.theguardian.com/the-dawn-of-digital-therapeutics/2025/jan/03/security-is-vital-how-should-new-health-innovations-protect-user-data-and-put-it-to-good-use

Fernández, A. L., & Marcopulos, B. A. (2008). A comparison of normative data for the Trail Making Test from several countries: Equivalence of norms and considerations for interpretation: Cognition and Neurosciences. *Scandinavian Journal of Psychology*, *49*(3), 239–246. https://doi.org/10.1111/j.1467-9450.2008.00637.x

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Franciotti, R., Nardini, D., Russo, M., Onofrj, M., & Sensi, S. L. (2023). Comparison of Machine Learning-based Approaches to Predict the Conversion to Alzheimer's Disease from Mild Cognitive Impairment. *Neuroscience*, *514*, 143–152. https://doi.org/10.1016/j.neuroscience.2023.01.029

González-Erena, P. V, Fernández-Guinea, S., & Kourtesis, P. (2025). Cognitive Assessment and Training in Extended Reality: Multimodal Systems, Clinical Utility, and Current Challenges. *Encyclopedia*, *5*(1), 8. https://doi.org/10.3390/encyclopedia5010008

Grammenos, G., Vrahatis, A. G., Vlamos, P., Palejev, D., & Exarchos, T. (2024). Predicting the Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using an Explainable AI Approach. *Information (Switzerland)*, *15*(5). https://doi.org/10.3390/info15050249

Guzmán, D. E., Rengifo, C. F., & García-Cena, C. E. (2024). Serious Games for Cognitive Rehabilitation in Older Adults: A Conceptual Framework. *Multimodal Technologies and Interaction*, *8*(8). https://doi.org/10.3390/mti8080064

Hammady, R., & Arnab, S. (2022). Serious Gaming for Behaviour Change: A Systematic Review. *Information (Switzerland)*, *13*(3). https://doi.org/10.3390/info13030142

Hao, J., Kwapong, W. R., Shen, T., Fu, H., Xu, Y., Lu, Q., Liu, S., Zhang, J., Liu, Y., Zhao, Y., Zheng, Y., Frangi, A. F., Zhang, S., Qi, H., & Zhao, Y. (2024). Early detection of dementia through retinal imaging and trustworthy AI. *Npj Digital Medicine*. https://doi.org/10.1038/s41746-024-01292-5

Hendrix, M., Bellamy-Wood, T., McKay, S., Bloom, V., & Dunwell, I. (2019). Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*, *11*(4), 320–327. https://doi.org/10.1109/TG.2018.2791019

Hildebrandt, H. (2020). Neuropsychological Tools for Dementia: Differential Diagnosis and Treatment. In *Neuropsychological Tools for Dementia: Differential Diagnosis and Treatment*. Academic Press. https://doi.org/10.1016/B978-0-12-821072-7.00010-2

Horan, W. P., Moore, R. C., Belanger, H. G., & Harvey, P. D. (2024). Utilizing Technology to Enhance the Ecological Validity of Cognitive and Functional Assessments in Schizophrenia: An Overview of the State-of-the-Art. *Schizophrenia Bulletin Open*, *5*(1). https://doi.org/10.1093/schizbullopen/sgae025

Hoseini, F. S., Khodadadi, M., & Khorambakht, A. (2022). The Effect of 2D and 3D Action Video Game Interventions on Executive Functions in Male Students. *Simulation and Gaming*, *53*(5), 405–422. https://doi.org/10.1177/10468781221110577

Hughes, R., Dalakas, M. C., Merkies, I., Latov, N., Léger, J. M., Nobile-Orazio, E., Sobue, G., Genge, A., Cornblath, D., Merschhemke, M., Ervin, C. M., Agoropoulou, C., Hartung, H. P., Day, T., Spies, J., Roberts, L., Van Damme, P., Van den Bergh, P. Y., Maertens de Noordhout, A., … Amato, A. A. (2018). Oral fingolimod for chronic inflammatory demyelinating polyradiculoneuropathy (FORCIDP Trial): a double-blind, multicentre, randomised controlled trial. *The Lancet Neurology*, *17*(8), 689–698. https://doi.org/10.1016/S1474-4422(18)30202-3

Ijaz, K., Tran, T. T. M., Kocaballi, A. B., Calvo, R. A., Berkovsky, S., & Ahmadpour, N. (2022). Design Considerations for Immersive Virtual Reality Applications for Older Adults: A Scoping Review. *Multimodal Technologies and Interaction*, *6*(7). https://doi.org/10.3390/mti6070060

ISC2. (2024). *The Ethical Dilemmas of AI in Cybersecurity*. The ISC2 Global Achievement Awards. https://www.isc2.org/Insights/2024/01/The-Ethical-Dilemmas-of AI-in-Cybersecurity

Ishaq, K., Rosdi, F., Zin, N. A. M., & Abid, A. (2022). Serious game design model for language learning in the cultural context. *Education and Information Technologies*, *27*(7), 9317–9355. https://doi.org/10.1007/s10639-022-10999-5

Jaeger, J. (2018). Digit symbol substitution test. *Journal of Clinical Psychopharmacology*, *38*(5), 513–519. https://doi.org/10.1097/JCP.0000000000000941

Javed, A. R., Saadia, A., Mughal, H., Gadekallu, T. R., Rizwan, M., Maddikunta, P. K. R., Mahmud, M., Liyanage, M., & Hussain, A. (2023). Artificial Intelligence for Cognitive Health Assessment: State-of-the-Art, Open Challenges and Future Directions. *Cognitive Computation*, *15*(6), 1767–1812. https://doi.org/10.1007/s12559-023-10153-4

Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*. https://doi.org/10.3389/fnagi.2019.00220

Kale, M., Wankhede, N., Pawar, R., Ballal, S., Kumawat, R., Goswami, M., Khalid, M., Taksande, B., Upaganlawar, A., Umekar, M., Kopalli, S. R., & Koppula, S. (2024). AI-driven innovations in Alzheimer's disease: Integrating early diagnosis, personalized treatment, and prognostic modelling. *Ageing Research Reviews*, *101*. https://doi.org/10.1016/j.arr.2024.102497

Khaleghi, A., Aghaei, Z., & Mahdavi, M. A. (2021). A gamification framework for cognitive assessment and cognitive training: Qualitative study. *JMIR Serious Games*, *9*(2). https://doi.org/10.2196/21900

Kirkham, R., Kooijman, L., Albertella, L., Myles, D., Yücel, M., & Rotaru, K. (2024). Immersive Virtual Reality-Based Methods for Assessing Executive Functioning: Systematic Review. *JMIR Serious Games*, *12*(1). https://doi.org/10.2196/50282

Koivisto, J., & Malik, A. (2021). Gamification for Older Adults: A Systematic Literature Review. *Gerontologist*, *61*(7), E345–E357. https://doi.org/10.1093/geront/gnaa047

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 Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0084-2

Kourtesis, P., & MacPherson, S. E. (2021). How immersive virtual reality methods may meet the criteria of the National Academy of Neuropsychology and American Academy of Clinical Neuropsychology: A software review of the Virtual Reality Everyday Assessment Lab (VR-EAL). *Computers in Human Behavior Reports, 4*, 100151. https://doi.org/10.1016/j.chbr.2021.100151

Krawczyk, D. (2021). *Virtual reality assessment effectively tests executive function in a real-world setting.* Center for BrainHealth. https://centerforbrainhealth.org/article/researchers-create-virtual-reality-cognitive-assessment

Kurbanova, M. M., Galayeva, A. A., Stefanovskaya, Y. V., Suvorkina, A. A., & Alikhanov, N. M. (2020). Modern methods for the diagnosis of cognitive impairment. *Russian Family Doctor, 24*(1), 35–44. https://doi.org/10.17816/RFD18986

Lesport, Q., Palmie, D., Öztosun, G., Kaminski, H. J., & Garbey, M. (2024). AI-Powered Telemedicine for Automatic Scoring of Neuromuscular Examinations. *Bioengineering*, *11*(9). https://doi.org/10.3390/bioengineering11090942

Leutner, F., Codreanu, S. C., Brink, S., & Bitsakis, T. (2023). Game based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.942662

Li, A., Li, J., Zhang, D., Wu, W., Zhao, J., & Qiang, Y. (2023). Synergy through integration of digital cognitive tests and wearable devices for mild cognitive impairment screening. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1183457

Li, C., Lin, R., Liu, Y., & Wei, Y. (2024). Evaluating Cognitive and Neuropsychological Assessments - A Comprehensive Review. *ArXiv Preprint ArXiv*, *2402.14655*. https://doi.org/10.48550/arXiv.2402.14655

Li, H., & Fan, Y. (2019). Early prediction of Alzheimer's disease dementia based on baseline hippocampal mri and 1-year follow-up cognitive measures using deep recurrent neural networks. *Proceedings - International Symposium on Biomedical Imaging*, *2019-April*, 368–371. https://doi.org/10.1109/ISBI.2019.8759397

Lickiewicz, J., Hughes, P. P., & Makara-Studzińska, M. (2020). Serious Games and Board Games Versus Cultural Changes. *Perspektywy Kultury*, *30*(3), 257–270. https://doi.org/10.35765/pk.2020.3003.17

Lin, M., Gong, P., Yang, T., Ye, J., Albin, R. L., & Dodge, H. H. (2018). Big Data Analytical Approaches to the NACC Dataset: Aiding Preclinical Trial Enrichment. *Alzheimer Disease and Associated Disorders*, *32*(1), 18–27. https://doi.org/10.1097/WAD.0000000000000228

Liszewski, A. (2024). *This head-tracking spinning chair could make VR less nauseating*. Theverge. https://www.theverge.com/2024/8/14/24220354/vr-meta-quest-roto-explorer-chair-virtual-reality

Liu, Y., Tan, W., Chen, C., Liu, C., Yang, J., & Zhang, Y. (2019). A Review of the Application of Virtual Reality Technology in the Diagnosis and Treatment of Cognitive Impairment. *Frontiers in Aging Neuroscience*, *11*. https://doi.org/10.3389/fnagi.2019.00280

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., … Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, *396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., Cooper, C., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Larson, E. B., Ritchie, K., Rockwood, K., Sampson, E. L., … Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet*, *390*(10113), 2673–2734. https://doi.org/10.1016/S0140-6736(17)31363-6

Loukou, E. (2025). *WATCH THIS How to predict your likelihood of getting dementia using just your fitness tracker*. The Sun. https://www.thesun.co.uk/health/33225699/predict-likelihood-dementia-fitness-tracker-steps-sleep/

Lu, L., Zhang, J., Xie, Y., Gao, F., Xu, S., Wu, X., & Ye, Z. (2020). Wearable health devices in health care: Narrative systematic review. *JMIR MHealth and UHealth*, *8*(11). https://doi.org/10.2196/18907

Lui, Lasse, F., Radhakrishnan, U., Chinello, F., & Koumaditis, K. (2025). The efficacy of adaptive training in immersive virtual reality for a fine motor skill task. *Virtual Reality*, *29:20*(1). https://doi.org/10.1007/s10055-024-01083-z

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games*, *4*(2). https://doi.org/10.2196/games.5888

Manly, J. J. (2008). Critical issues in cultural neuropsychology: Profit from diversity. *Neuropsychology Review*, *18*(3 SPEC. ISS.), 179–183. https://doi.org/10.1007/s11065-008-9068-8

Marin-Morales, J., Carrasco-Ribelles, L. A., Alcaniz, M., & Giglioli, I. A. C. (2021). Applying machine learning to a virtual serious game for neuropsychological assessment. *IEEE Global Engineering Education Conference, EDUCON*, *2021-April*, 946–949. https://doi.org/10.1109/EDUCON46332.2021.9454138

McKenzie, B. (2025a). *New AI tool predicts brain decline years before symptoms appear — how it could help fight dementia*. Nypost. https://nypost.com/2025/03/08/health/new-ai-tool-predicts-brain-decline-before-symptoms-appear/

McKenzie, B. (2025b). *'Smart pajamas' can monitor sleep apnea and other disorders at home — and send the results to your smartphone*. Nypost. https://nypost.com/2025/02/21/health/smart-pajamas-can-monitor-sleep-disorders-at-home/

Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, *10*(4). https://doi.org/10.1016/j.heliyon.2024.e26297

Merriman, N. A., Roudaia, E., Ondřej, J., Romagnoli, M., Orvieto, I., O'Sullivan, C., & Newell, F. N. (2022). "CityQuest," A Custom-Designed Serious Game, Enhances Spatial Memory Performance in Older Adults. *Frontiers in Aging Neuroscience*, *14*. https://doi.org/10.3389/fnagi.2022.806418

Migal, A., & Nazarov, V. (2024). *AI in Telehealth: Revolutionizing Healthcare Delivery to Every Patient's Home*. TATEEDA.Com. https://tateeda.com/blog/ai-in-telemedicine-use-cases

Mireshghallah, F., Taram, M., Vepakomma, P., Singh, A., Raskar, R., & Esmaeilzadeh, H. (2020). *Privacy in Deep Learning: A Survey*. http://arxiv.org/abs/2004.12254

Mittal, S., Thakral, K., Singh, R., Vatsa, M., Glaser, T., Ferrer, C. C., & Hassner, T. (2023). *On Responsible Machine Learning Datasets with Fairness, Privacy, and Regulatory Norms*. http://arxiv.org/abs/2310.15848

Morán, J. F. O., Blas Pagador, J., Preciado, V. G., Moyano-Cuevas, J. L., Domínguez, T. R., Muñoz, M. S., & Sánchez Margallo, F. M. (2024). A Serious Game for Cognitive Stimulation of Older People With Mild Cognitive Impairment: Design and Pilot Usability Study. *JMIR Aging*, *7*, 1–14. https://doi.org/10.2196/41437

Mouatt, B., Smith, A. E., Mellow, M. L., Parfitt, G., Smith, R. T., & Stanton, T. R. (2020). The Use of Virtual Reality to Influence Motivation, Affect, Enjoyment, and Engagement During Exercise: A Scoping Review. *Frontiers in Virtual Reality*, *1*. https://doi.org/10.3389/frvir.2020.564664

Moulaei, K., Sharifi, H., Bahaadinbeigy, K., & Dinari, F. (2024). Efficacy of virtual reality-based training programs and games on the improvement of cognitive disorders in patients: a systematic review and meta-analysis. *BMC Psychiatry*, *24*(1). https://doi.org/10.1186/s12888-024-05563-z

Mullin, E. (2024). *I Tried These Brain-Tracking Headphones That Claim to Improve Focus*. WIRED. https://www.wired.com/story/this-brain-tracking-device-wants-to-help-you-work-smarter/

Musto, H., Stamate, D., Logofatu, D., & Stahl, D. (2024). Predicting Deterioration in Mild Cognitive Impairment with Survival Transformers, Extreme Gradient Boosting and Cox Proportional Hazard Modelling. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *15023 LNCS*, 359–372. https://doi.org/10.1007/978-3-031-72353-7_26

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nguyen, L., Murphy, K., & Andrews, G. (2025). Design and development of a gamified cognitive training program targeting executive functions for older adults. *Entertainment Computing*, *52*. https://doi.org/10.1016/j.entcom.2024.100892

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10). https://doi.org/10.1016/j.patter.2021.100347

O'meara, A., Cassarino, M., Bolger, A., & Setti, A. (2020). Virtual reality nature exposure and test anxiety. *Multimodal Technologies and Interaction*, *4*(4), 1–14. https://doi.org/10.3390/mti4040075

Obuchi, S. P., Kojima, M., Suzuki, H., Garbalosa, J. C., Imamura, K., Ihara, K., Hirano, H., Sasai, H., Fujiwara, Y., & Kawai, H. (2024). Artificial intelligence detection of cognitive impairment in older adults during walking. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *16*(3). https://doi.org/10.1002/dad2.70012

Othman, N. I., Mohamed, H., & Mat Zin, N. A. (2023). Serious Games Accessibility Design Model for Low-Vision Children. *Advances in Human-Computer Interaction*, *2023*. https://doi.org/10.1155/2023/9528294

OVIC. (2021). *Artificial Intelligence – Understanding Privacy Obligations*. Office of the Victorian Information Commissioner. https://ovic.vic.gov.au/privacy/resources-for-organisations/artificial-intelligence-understanding-privacy-obligations/

Özgen, D. S., Afacan, Y., & Sürer, E. (2021). Usability of virtual reality for basic design education: a comparative study with paper-based design. *International Journal of Technology and Design Education*, *31*(2), 357–377. https://doi.org/10.1007/s10798-019-09554-0

Parsons, T. D. (2011). Neuropsychological assessment using virtual environments: Enhanced assessment technology for improved ecological validity. In *Studies in Computational Intelligence* (Vol. 337). https://doi.org/10.1007/978-3-642-17824-5_13

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*(DEC). https://doi.org/10.3389/fnhum.2015.00660

Pedroli, E., Serino, S., Cipresso, P., Pallavicini, F., & Riva, G. (2015). Assessment and rehabilitation of neglect using virtual reality: A systematic review. *Frontiers in Behavioral Neuroscience*, *9*(AUGUST). https://doi.org/10.3389/fnbeh.2015.00226

Piau, A., Wild, K., Mattek, N., & Kaye, J. (2019). Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: Systematic review. *Journal of Medical Internet Research*, *21*(8). https://doi.org/10.2196/12785

Pistono, A. M. A. de A., dos Santos, A. M. P., Baptista, R. J. V., & Mamede, H. S. (2024). Framework for adaptive serious games. *Computer Applications in Engineering Education*, *32*(4). https://doi.org/10.1002/cae.22731

Pratviel, Y., Bouny, P., & Deschodt-Arsac, V. (2024). Immersion in a relaxing virtual reality environment is associated with similar effects on stress and anxiety as heart rate variability biofeedback.

Frontiers in Virtual Reality, *5*. https://doi.org/10.3389/frvir.2024.1358981

Rajasegeran, D. D., Liu, K., Sheng, Y., Loh, C. S., Choh, A. C. L., Teo, K. Y., Fan, P. E. M., Tan, M. Y., Aloweni, F., & Ang, S. Y. (2023). Potential of serious games as a competency assessment tool for acute care nurses on the blood transfusion procedure. *International Journal of Digital Health*. https://doi.org/10.1097/jh9.0000000000000006

Redlinger, E., Glas, B., & Rong, Y. (2022). Impact of Visual Game-Like Features on Cognitive Performance in a Virtual Reality Working Memory Task: Within-Subjects Experiment. *JMIR Serious Games*, *10*(2). https://doi.org/10.2196/35295

Reidy, L., Chan, D., Nduka, C., & Gunes, H. (2020). Facial Electromyography-based Adaptive Virtual Reality Gaming for Cognitive Training. *ICMI 2020 - Proceedings of the 2020 International Conference on Multimodal Interaction*, 174–183. https://doi.org/10.1145/3382507.3418845

Romero-Ayuso, D., del Pino-González, A., Torres-Jiménez, A., Juan-González, J., Celdrán, F. J., Franchella, M. C., Ortega-López, N., Triviño-Juárez, J. M., Garach-Gómez, A., Arrabal-Fernández, L., Medina-Martínez, I., & González, P. (2024). Enhancing Ecological Validity: Virtual Reality Assessment of Executive Functioning in Children and Adolescents with ADHD. *Children*, *11*(8). https://doi.org/10.3390/children11080986

Rykov, Y. G., Patterson, M. D., Gangwar, B. A., Jabar, S. B., Leonardo, J., Ng, K. P., & Kandiah, N. (2024). Predicting cognitive scores from wearable-based digital physiological features using machine learning: data from a clinical trial in mild cognitive impairment. *BMC Medicine*, *22*(1). https://doi.org/10.1186/s12916-024-03252-y

Sakal, C., Li, T., Li, J., & Li, X. (2023). *Assessing cognitive function among older adults using machine learning and wearable device data: A feasibility study* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2309.07133

Seifert, A., & Schlomann, A. (2021). The Use of Virtual and Augmented Reality by Older Adults: Potentials and Challenges. *Frontiers in Virtual Reality*, *2*. https://doi.org/10.3389/frvir.2021.639718

Shahmoradi, L., Mohammadian, F., & Rahmani Katigari, M. (2022). A Systematic Review on Serious Games in Attention Rehabilitation and Their Effects. *Behavioural Neurology*, *2022*. https://doi.org/10.1155/2022/2017975

Shankar, R., Bundele, A., & Mukhopadhyay, A. (2025). Natural language processing of electronic health records for early detection of cognitive decline: a systematic review. *Npj Digital Medicine*, *8*(133). https://doi.org/10.1038/s41746-025-01527-z

Sharma, S., Rawal, R., & Shah, D. (2023). Addressing the challenges of AI-based telemedicine: Best practices and lessons learned. *Journal of Education and Health Promotion*, *12*(1). https://doi.org/10.4103/jehp.jehp_402_23

Silverberg, N. B., Ryan, L. M., Carrillo, M. C., Sperling, R., Petersen, R. C., Posner, H. B., Snyder, P. J., Hilsabeck, R., Gallagher, M., Raber, J., Rizzo, A., Possin, K., King, J., Kaye, J., Ott, B. R., Albert, M. S., Wagster, M. V., Schinka, J. A., Cullum, C. M., … Ferman, T. J. (2011). Assessment of cognition in early dementia. *Alzheimer's & Dementia*, *7*(3). https://doi.org/10.1016/j.jalz.2011.05.001

Sokołowska, B. (2023). Impact of Virtual Reality Cognitive and Motor Exercises on Brain Health. *International Journal of Environmental Research and Public Health*, *20*(5). https://doi.org/10.3390/ijerph20054150

Song, H., Yi, D. J., & Park, H. J. (2020). Validation of a mobile game-based assessment of cognitive control among children and adolescents. *PLoS ONE*, *15*(3). https://doi.org/10.1371/journal.pone.0230498

Streicher, A., & Smeddinck, J. D. (2016). Personalized and adaptive serious games. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9970 LNCS*, 332–377. https://doi.org/10.1007/978-3-319-46152-6_14

Swahn, M. H., Gittner, K. B., Lyons, M. J., Nielsen, K., Mobley, K., Culbreth, R., Palmier, J., Johnson, N. E., Matte, M., & Nabulya, A. (2024). Advancing mHealth Research in Low-Resource Settings:

Young Women's Insights and Implementation Challenges with Wearable Smartwatch Devices in Uganda. *Sensors*, *24*(17). https://doi.org/10.3390/s24175591

Tan, N. C., Lim, J. E., Sultana, R., Quah, J. H. M., & Wong, W. T. (2024). A virtual reality cognitive screening tool based on the six cognitive domains. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *16*(4). https://doi.org/10.1002/dad2.70030

Tan, W. Y., Hargreaves, C., Chen, C., & Hilal, S. (2023). A Machine Learning Approach for Early Diagnosis of Cognitive Impairment Using Population-Based Data. *Journal of Alzheimer's Disease*, *91*(1), 449–461. https://doi.org/10.3233/JAD-220776

Tascedda, S., Sarti, P., Rivi, V., Guerrera, C. S., Platania, G. A., Santagati, M., Caraci, F., & Blom, J. M. C. (2024). Advanced AI techniques for classifying Alzheimer's disease and mild cognitive impairment. *Frontiers in Aging Neuroscience*, *16*. https://doi.org/10.3389/fnagi.2024.1488050

Thongthip, P., Intawong, K., Chernbumroong, S., Worragin, P., Tongpaeng, Y., & Puritat, K. (2024). Design and analysis of personalized serious games for information literacy: catering to introverted and extraverted individuals through game elements. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-03172-5

Timaná, L. C. R., García, J. F. C., Filho, T. B., González, A. A. O., Monsalve, N. R. H., & Jimenez, N. J. V. (2024). Use of Serious Games in Interventions of Executive Functions in Neurodiverse Children: Systematic Review. *JMIR Serious Games*, *12*. https://doi.org/10.2196/59053

Tsoi, K. K. F., Jia, P., Dowling, N. M., Titiner, J. R., Wagner, M., Capuano, A. W., & Donohue, M. C. (2023). Applications of artificial intelligence in dementia research. *Cambridge Prisms: Precision Medicine*, *1*. https://doi.org/10.1017/pcm.2022.10

Urbina, S., Kaufman, A. ., & Kaufman, N. . (2014). *Essentials of Psychological Testing*. John Wiley & Sons. http://pustaka.unp.ac.id/file/abstrak_kki/EBOOKS/Essentials of Psychological Testing.pdf

Vercruyssen, A., Schirmer, W., & Mortelmans, D. (2023). How "basic" is basic digital literacy for older adults? Insights from digital skills instructors. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1231701

Wang, J., Zhou, Z., Cheng, S., Zhou, L., Sun, X., Song, Z., Wu, Z., Lu, J., Qin, Y., & Wang, Y. (2024). Dual-task turn velocity – a novel digital biomarker for mild cognitive impairment and dementia. *Frontiers in Aging Neuroscience*, *16*. https://doi.org/10.3389/fnagi.2024.1304265

Watson, N. L., Rosano, C., Boudreau, R. M., Simonsick, E. M., Ferrucci, L., Sutton-Tyrrell, K., Hardy, S. E., Atkinson, H. H., Yaffe, K., Satterfield, S., Harris, T. B., & Newman, A. B. (2010). Executive function, memory, and gait speed decline in well-functioning older adults. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *65 A*(10), 1093–1100. https://doi.org/10.1093/gerona/glq111

White, Z. (2023). Re-envisioning Digital Equity and Connection Literacy for Older Adults. *North Carolina Medical Journal*, *84*(2), 110–113. https://doi.org/10.18043/001c.73013

Wilczyńska, D., Walczak-Kozłowska, T., Alarcón, D., Arenilla, M. J., Jaenes, J. C., Hejła, M., Lipowski, M., Nestorowicz, J., & Olszewski, H. (2025). The Role of Immersive Experience in Anxiety Reduction: Evidence from Virtual Reality Sessions. *Brain Sciences*, *15*(1). https://doi.org/10.3390/brainsci15010014

Winblad, B., Amouyel, P., Andrieu, S., Ballard, C., Brayne, C., Brodaty, H., Cedazo-Minguez, A., Dubois, B., Edvardsson, D., Feldman, H., Fratiglioni, L., Frisoni, G. B., Gauthier, S., Georges, J., Graff, C., Iqbal, K., Jessen, F., Johansson, G., Jönsson, L., … Zetterberg, H. (2016). Defeating Alzheimer's disease and other dementias: A priority for European science and society. *The Lancet Neurology*, *15*(5), 455–532. https://doi.org/10.1016/S1474-4422(16)00062-4

Xie, X., Shi, R., Yu, H., Wan, X., Liu, T., Duan, D., Li, D., & Wen, D. (2024). Executive function rehabilitation and evaluation based on brain-computer interface and virtual reality: our opinion. *Frontiers in Neuroscience*, *18*. https://doi.org/10.3389/fnins.2024.1377097

Xue, C., Kowshik, S. S., Lteif, D., Puducheri, S., Jasodanand, V. H., Zhou, O. T., Walia, A. S., Guney, O. B., Zhang, J. D., Pham, S. T., Kaliaev, A., Andreu-Arasa, V. C., Dwyer, B. C., Farris, C. W., Hao,

H., Kedar, S., Mian, A. Z., Murman, D. L., O'Shea, S. A., … Kolachalama, V. B. (2024). AI-based differential diagnosis of dementia etiologies on multimodal data. *Nature Medicine*. https://doi.org/10.1038/s41591-024-03118-z

Yang, J., Soltan, A. A. S., Eyre, D. W., & Clifton, D. A. (2023). Algorithmic fairness and bias mitigation for clinical machine learning with deep reinforcement learning. *Nature Machine Intelligence*, *5*(8), 884–894. https://doi.org/10.1038/s42256-023-00697-3

Zhou, Y., Zhao, Y., Xiang, Z., Yan, Z., Shu, L., Xu, X., Zhang, L., & Tian, X. (2023). A dual-task-embedded virtual reality system for intelligent quantitative assessment of cognitive processing speed. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1158650

Zorowitz, S., & Niv, Y. (2023). Improving the Reliability of Cognitive Task Measures: A Narrative Review. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(8), 789–797. https://doi.org/10.1016/j.bpsc.2023.02.004

Zygouris, S., Ntovas, K., Giakoumis, D., Votis, K., Doumpoulakis, S., Segkouli, S., Karagiannidis, C., Tzovaras, D., & Tsolaki, M. (2017). A Preliminary Study on the Feasibility of Using a Virtual Reality Cognitive Training Application for Remote Detection of Mild Cognitive Impairment. *Journal of Alzheimer's Disease*, *56*(2), 619–627. https://doi.org/10.3233/JAD-160518

Received: September 7, 2025
Revision received: December 28, 2025
Accepted: March 30, 2026

Author Contributions

Nazyar Khamenehei contributed to the conceptualization and design of the study, development of the methodology, and the execution of the investigation. They were responsible for data analysis and interpretation of the results. Nazyar also prepared the original draft of the manuscript, and contributed to writing, reviewing, and editing the final version of the paper.

Lyudmila V. Tokarskaya provided supervision and oversight of the study, contributed to project preparation and administration, and participated in the investigation. She also reviewed and edited the manuscript to ensure accuracy and clarity.

Author Details

Nazyar Khamenehei – PhD student, General Psychology, Personality Psychology, History of Psychology, Research_Engineer, Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, Russia, ORCID ID: https://orcid.org/0000-0003-0190-5007, email: Nazyarkh@Gmail.com

Lyudmila V. Tokarskaya – Associate Professor, Deputy Director, Candidate of Psychological Sciences, Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, Russia, ORCID ID: https://orcid.org/0000-0002-2385-9227, email: liydmil@mail.ru

Conflict of Interest Information

The authors have no conflicts of interest to declare.