International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

Psychometric properties of digital cognitive assessments across the cognitive impairment continuum: an umbrella review protocol

INPLASY202660108
doi: 10.37766/inplasy2026.6.0108
Received: 22 June 2026
Published: 22 June 2026

Panesar, JR; Wilschut, TJ; van Rijn, H

Corresponding author:
Joshua Panesar

j.r.panesar@student.rug.nl

Author Affiliation:
University of Groningen, Groningen, The Netherlands.

ADMINISTRATIVE INFORMATION

Support - NWO Take-off phase 2.

Review Stage at time of this submission - Preliminary searches.

Conflicts of interest - JP, TW, and HvR are based at the University of Groningen and are also associated with Precision Cognition Labs, a commercial partner involved in this project that develops SGMA, a digital memory test, for which the authors have an interest in validation. This interest is explicitly declared. As mitigation, the eligibility criteria, outcomes, and appraisal approach are pre-specified in this prospectively registered protocol, before screening begins; should the tool be covered by any review included in this umbrella review, it is screened, extracted, and appraised under the same pre-specified procedures as every other tool, independently and in duplicate, with no preferential weighting or interpretation; and all such decisions are documented so that they can be audited. All other authors declare no competing interests.

INPLASY registration number: INPLASY202660108

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 22 June 2026 and was last updated on 22 June 2026.

INTRODUCTION

Review question / Objective This umbrella review aims to systematically map and synthesise evidence from existing review articles (systematic reviews, meta-analyses, scoping reviews, umbrella reviews, reviews-of-reviews, narrative reviews, and literature reviews) on the psychometric properties of digital, home-administrable cognitive assessment tools that include at least one memory component, across the cognitive impairment continuum, with non-Alzheimer's dementia included as a parallel population column rather than a stage of the continuum (see Item 10).
The review will address the following questions:

- Where does the existing review evidence sit methodologically across the continuum? That is, which continuum stages are well covered by methodologically high-quality reviews, and where are the methodological coverage gaps?
- Which digital, home-administrable cognitive assessment tools with a memory component have been evaluated in the review literature with reported psychometric properties, and at which stage(s) of the continuum?
- What psychometric evidence (reliability, convergent validity, and diagnostic accuracy) has been reported for each tool, at which stage(s)?
- Which tools have the strongest psychometric support for use in a given stage of the continuum, and where are the most important evidence gaps at the tool level?

The principal outputs are evidence maps that display, across the continuum, the methodological coverage of the review evidence and the reliability, convergent validity, and diagnostic accuracy reported for individual tools (Item 22).

Rationale Digital, home-administrable cognitive assessments have proliferated rapidly over the past decade as a solution to the scalability and access limitations of in-clinic neuropsychological testing, and as a possibility to detect subtle cognitive change earlier in the cognitive impairment continuum than conventional instruments often allow. As the number of available tools has grown, so has the number of reviews evaluating them: Several systematic, scoping, and narrative reviews address overlapping sets of digital cognitive tests, populations, and measurement properties. These reviews differ in how they define "digital," which stages of the continuum they cover, which psychometric properties they report, and which reference standards they accept. In addition, they frequently include the same primary studies. The result is a fragmented evidence base of diagnostic test accuracy studies, which makes it difficult to determine, for any given tool, what is actually known about its reliability, validity, and diagnostic accuracy, and at which stage of the continuum that evidence applies.

To the best of our knowledge, no umbrella review has yet synthesised the systematic-review-level evidence on the psychometric properties of these tools across the continuum. Existing reviews are predominantly organised by tool or by technology type rather than by stage of the continuum, and none maps the psychometric evidence for individual tools against the stage of disease in which each tool has been validated. The existing reviews on digital cognitive assessment across the cognitive impairment continuum mostly focus on a segment of the continuum, a single psychometric property, and/or review a specific subgroup of the digital cognitive assessments. Moreover, no review has established the methodological quality of the underlying review literature. Where multiple reviews report on the same tool, their estimates have not been reconciled, and the degree of primary-study overlap between them has not been quantified.

An umbrella review is the appropriate design for this problem because the unit requiring synthesis is the review rather than the primary study. By appraising and synthesising the existing reviews rather than re-extracting primary data, this review aims to establish, for each digital cognitive tool with a memory component, what psychometric evidence has been reported in the review literature and at which stage of the continuum, to identify which tools have the strongest evidence for a given stage, and to make explicit where the review literature is methodologically weak, internally inconsistent, or absent. This review, therefore, contributes both: a clinically usable map of which tools are adequately supported for which populations, and a methodological map of where the field's synthesised evidence is missing and where new primary research or new reviews are most needed.

Condition being studied Cognitive function across the cognitive impairment continuum, from unimpaired cognition through to dementia.

The four stages are ordered by increasing cognitive impairment attributable to Alzheimer's disease, while the non-AD dementia column sits outside that ordering as a parallel population rather than a more severe stage. The continuum is operationalised as categorical stages following the NIA-AA framework (1).

1. Healthy, cognitively unimpaired: cognitively unimpaired older adults, healthy aging, healthy controls, and preclinical AD (biomarker-positive but cognitively unimpaired, NIA-AA clinical stage 1)
2. Subjective Cognitive Decline: subjective complaints with normal objective performance, including subtle cognitive change not meeting MCI criteria
3. Mild Cognitive Impairment: objective impairment not meeting dementia criteria, including MCI due to AD and prodromal AD
4. Alzheimer's dementia: including early-stage AD
5. non-Alzheimer's dementia: vascular, Lewy body, frontotemporal, and mixed dementia included as a parallel population column, not as a more severe stage of the continuum. Whether psychometric properties observed in non-AD populations differ from those reported in AD is treated descriptively in the discussion. No a priori expectation of differential performance is built into the synthesis.

Preclinical AD is clustered under the cognitively unimpaired stage rather than with MCI. The NIA-AA definition denotes biomarker abnormalities in the absence of objective cognitive impairment. In contrast, "prodromal AD" is equivalent to MCI due to AD and is mapped to the MCI stage. Stage assignment follows the operationalised sample described in each review, not its nominal label in the title, per the rule in Item 20.

METHODS

Search strategy A five-block Boolean search strategy was developed combining (Block 1) digital technologies AND (Block 2) cognitive assessment tools AND (Block 3) psychometric properties AND (Block 4) cognitive impairment continuum

populations AND (Block 5) review-type publication filter. Each block combines free-text title/abstract terms (with truncation and phrase searching) and database-specific controlled vocabulary. Filters applied across all databases: English-language, publication years 2006–2026.
Searches were run on the following dates with the following result counts:
– PubMed – 213 records (4 May 2026)
– Scopus – 156 records (5 May 2026)
– MEDLINE via EBSCOhost – 141 records (5 May 2026)
– APA PsycINFO via EBSCOhost – 89 records (5 May 2026)
– Web of Science Core Collection – 140 records (6 May 2026)
– Cochrane Library – 79 records (6 May 2026)
The complete Boolean search string for each database, with its interface, controlled-vocabulary mapping, applied filters, exact run time, and result count, has been recorded in full and will be reported with the published review in accordance with PRISMA-S.

Participant or population Adults aged 18 years or older across the cognitive impairment continuum: cognitively unimpaired/healthy aging older adults, individuals with subjective cognitive decline (SCD), individuals with mild cognitive impairment (MCI), prodromal Alzheimer's disease, early-stage dementia, clinical Alzheimer's dementia, mixed dementia, vascular dementia, Lewy body dementia, or frontotemporal dementia.

Reviews focused exclusively on populations outside this continuum, for example, schizophrenia, traumatic brain injury, Parkinson's disease, Huntington's disease, stroke without comorbid cognitive decline, ADHD, autism, or pediatric populations, are not eligible.

Mixed reviews (those covering both continuum and non-AD populations, or both digital and non-digital tools) are retained at the title/abstract stage. Whether continuum-specific and digital-tool-specific psychometric data can be separately extracted is assessed at the full-text screening stage rather than at title/abstract screening.

Intervention This is a psychometric-properties umbrella review rather than an intervention umbrella review. The "index" of interest is digital cognitive assessment tools that satisfy both of the following:
(a) Home-administrable. Designed for, or validated in, administration outside the clinical setting. This includes fully unsupervised home-based delivery and remotely supervised delivery (e.g., via videoconference).
Tools that require in-person clinical administration and have no remote or home-based component are not eligible.
(b) At least one memory component. Any task measuring encoding, retention, or retrieval of information, including delayed recall, recognition, associative/paired-associate learning, episodic memory, spatial memory, working memory, and prospective memory tasks.
Tools assessing only attention, processing speed, or other non-memory cognitive domains do not qualify.
"Digital" includes computerised tests, web-based and tablet-based batteries, smartphone applications, ecological momentary assessment, telephone-administered cognitive batteries, tele-neuropsychology platforms, virtual reality, gamified or serious-game tasks, and any other electronically delivered cognitive assessment. Paper-and-pencil tests and in-person clinic-administered tests with no digital component are excluded.

Comparator Variable depending on the psychometric property under evaluation, and not a primary structuring axis of this review:
• For diagnostic accuracy outcomes (sensitivity, specificity, AUC): the reference standard used in the underlying primary studies, typically a clinical diagnosis of MCI or AD (NIA-AA, DSM-5, or equivalent criteria) and/or a comprehensive neuropsychological battery.
• For criterion or convergent validity: established cognitive assessments such as the MoCA, MMSE, or established neuropsychological batteries.
• For test–retest reliability: no external comparator (a within-instrument property; see Item 18).

Study designs to be included Review-type publications: systematic reviews, meta-analyses, scoping reviews, umbrella reviews, reviews-of-reviews, narrative reviews, and literature reviews. Excluded: primary empirical studies, case reports, editorials, letters, conference abstracts, commentaries, book chapters, and review protocols without results reported.

Eligibility criteria Eligibility is specified against the PICOS elements. Because this is an umbrella review of psychometric properties rather than an intervention review, the Comparator element is not used as an eligibility criterion (see note below). Each criterion is applied at one of two screening stages: title/abstract (T/A) or full text (FT), as indicated. The stage is determined by whether the judgment can be made from the title and abstract alone or requires the full text (Item 20). Criteria

requiring assessment of whether continuum-specific or digital-tool-specific psychometric data are separately extractable are applied at the full text, because that judgment requires the complete paper.
Population (Item 12). Include reviews covering at least one of (a) healthy individuals or (b) the cognitive impairment continuum, even partially, healthy aging / cognitively unimpaired older adults, SCD, MCI, prodromal/preclinical AD, AD dementia, or non-AD dementia (vascular, Lewy body, frontotemporal, mixed). Exclude reviews that focus exclusively on non-AD neurodegenerative conditions and cover neither healthy individuals nor the cognitive impairment continuum (e.g., Parkinson's disease, Huntington's disease, multiple sclerosis without dementia). [T/A] Mixed-population reviews are retained at the title/abstract. The separability of continuum data is assessed in the full text. [FT]
Intervention / Index (Item 13). Include reviews covering at least one digital cognitive assessment tool that is (a) home-administrable, designed for, or validated in, fully unsupervised home-based or remotely supervised (e.g., videoconference) delivery. Exclude reviews in which all reviewed tools require in-person clinical administration with no remote or home-based component described or validated. [T/A] Mixed digital/non-digital reviews are retained at the title/abstract. Separability of digital-tool-specific data is assessed at the full text. [FT] A second requirement, that at least one digital tool includes a memory component (any task measuring encoding, retention, or retrieval of information, including delayed recall, recognition, and associative memory, while pure attention, processing speed, or other non-memory tasks do not qualify) is confirmed at full text. [FT]
Comparator (Item 14). This is not a formal eligibility criterion. The comparator varies for each psychometric property (reference standard for diagnostic accuracy, reference instrument for convergent validity, none for reliability) and is recorded during data extraction rather than used to include or exclude reviews. It is reported here for completeness only.
Outcome (Item 18). Include reviews reporting at least one psychometric property for at least one digital tool, specifically, reliability, validity, sensitivity, specificity, diagnostic accuracy, ICC, AUC, or comparable measurement-quality indicators. A review reporting no digital-tool-specific psychometric property is included only where continuum and digital-tool psychometric data are separately extractable and analysable. It is excluded where all psychometric results are pooled inseparably with non-digital tools and no digital-tool-specific figure, subgroup analysis, or narrative estimate can be identified. Reviews reporting only usability, feasibility, acceptability, or tool descriptions are excluded. [FT]
Study design (Item 15). Include review-type publications: systematic reviews, meta-analyses, scoping reviews, umbrella reviews, reviews-of-reviews, narrative reviews, and literature reviews. Exclude primary empirical studies, case reports, editorials, letters, conference abstracts, commentaries, and book chapters. [T/A] Exclude review protocols with no results reported, and reviews whose full text cannot be obtained after a reasonable effort. [FT]
Additional criteria not defined by PICOS (justified per INPLASY Item 16 guidance).
Language: English-language publications only. This restriction is applied for feasibility (the review team's working language) and is acknowledged as a potential source of language bias in the limitations.
Time frame: Publication years 2006–2026. The lower bound reflects the emergence of the modern computerised and home-based cognitive-assessment literature, as reviews of tools predating this period are unlikely to bear on currently deployable digital tools. The window is applied at the search stage and is not used as a post hoc exclusion.
Country/setting: No restriction.

Information sources Six bibliographic databases:
– PubMed
– Scopus
– MEDLINE via EBSCOhost
– APA PsycINFO via EBSCOhost
– Cochrane Library
– Web of Science Core Collection
Database access is via the University of Groningen RUG proxy.
Search reporting will follow PRISMA-S.

Main outcome(s) The main outcomes are the psychometric properties reported for the included digital cognitive assessment tools, corresponding to the tool-level evidence maps (Item 22):
Reliability: primarily test–retest reliability, recorded as the intraclass correlation coefficient (ICC) or test–retest correlation (r) as reported in the included reviews. Where an ICC is reported, its form (model and type) is recorded where stated and flagged 'unspecified' otherwise (2). Internal consistency (Cronbach's α, McDonald's ω) and inter-rater/intra-rater reliability are extracted where reported and presented in the supplementary tables.
Diagnostic accuracy: sensitivity and specificity (as a pair) against a clinical reference standard, and/or area under the receiver-operating-characteristic

curve (AUC). Reference standards include a clinical diagnosis of MCI or AD (NIA-AA, DSM-5, Petersen, or equivalent criteria), a full neuropsychological battery, or a biomarker-supported diagnosis.
Convergent validity: Pearson or Spearman correlation (r) between the digital tool's score and an established reference instrument, such as a global cognitive screener (the MoCA or MMSE) or a memory subtest of an established neuropsychological battery.

Benchmarks for what constitutes adequate reliability, validity, and diagnostic accuracy will be informed by established psychometric and diagnostic guidance (e.g., COSMIN/Terwee (3,4); consensus diagnostic-accuracy criteria (5); AUC interpretation (6)) and are finalised at the analysis stage. Their application to the maps is described in Item 22.

Additional outcome(s) Beyond the psychometric outcomes specified in Item 18, the following descriptive characteristics will be extracted at two levels (further detailed in Item 20) and will feed the evidence maps and accompanying narrative synthesis (Item 22):
At the review level: citation, review type, databases searched, number of included primary studies, continuum stages covered, tools covered, psychometric properties reported, reference standards used, and ROBIS risk-of-bias rating.
At the tool level: tool name (with cross-review tool-identity harmonisation, see Item 20), continuum stage of the population in which the tool was evaluated, sample size, cognitive domain(s) assessed (with particular attention to which memory subprocess is targeted, such as delayed recall, recognition, associative learning, spatial, working, or prospective memory), delivery mode (fully unsupervised home versus remotely supervised), device platform (web/computer, tablet, smartphone, telephone, VR), and administration time. Additional psychometric properties reported in the source review beyond the three primary outcomes (e.g., internal consistency, inter-rater reliability, structural validity, and responsiveness) are also extracted at the tool level and reported in the supplementary table.
The review-level data populate the review-level methodological-coverage map; the tool-level psychometric data populate the tool-level psychometric-property maps and the descriptive supplementary tables (Item 22).

Data management Reference management and deduplication. All records from the six databases are imported into Rayyan. Duplicate candidate pairs are resolved manually before screening, and only confirmed duplicates are removed; the number of confirmed duplicates and the unique-record count entering screening are reported in the PRISMA 2020 flow diagram.
Pilot screening. Approximately 30 abstracts are independently screened by all three reviewers before full title/abstract screening, to assess inter-rater agreement and refine the operationalisation of the criteria.
Title/abstract screening. Three independent reviewers (the lead author and two supervisors as co-screeners) screen all records in Rayyan, blinded to each other's decisions. Conflicts are resolved through discussion and persistent disagreements through adjudication.
Full-text screening. Dual screening by the same team; exclusions are documented with reasons (per the Stage-2 eligibility criteria in Item 16) for the PRISMA 2020 flow diagram.

Data extraction. A two-level extraction template is piloted on 2–3 anchor reviews and refined before full extraction. At the review level (one row per review): citation, review type, databases searched, number of primary studies, population and continuum stages covered, tools covered, psychometric outcomes reported, reference standards used, and ROBIS rating. At the tool level (one row per tool per stage per review): the harmonised tool name, the continuum stage of evaluation, sample size, the psychometric coefficients reported (test–retest ICC or r, sensitivity, specificity, AUC, convergent validity r, and the reference instrument), any additional psychometric properties, and whether digital-specific data are separately extractable in mixed reviews.
Stage attribution. Each extracted coefficient is attributed to a continuum stage using the stage stated in the review or, where unstated, the stage fixed by the review's inclusion criteria; coefficients that cannot be attributed to a single stage are recorded but handled conservatively in the synthesis (Item 22).
Tool-identity harmonisation. The same tool is often referred to by different names across reviews. A tool-identity list is built during pilot extraction and updated throughout the process, with the lead author and one co-reviewer independently classifying each new name and resolving disagreements through discussion. Decisions about which reviews contribute to which maps, and how discrepant estimates are reconciled, are described in Item 22.
Software. Rayyan (screening and deduplication), a spreadsheet-based extraction template, ROBIS, CCA calculation in a spreadsheet or R, and figure preparation in R or equivalent.

Quality assessment / Risk of bias analysis The risk of bias of each included review is appraised using ROBIS (7) as the only risk-of-bias tool. ROBIS is applied in its three phases: assessing relevance to the review question; identifying concerns across four domains (study eligibility criteria; identification and selection of studies; data collection and study appraisal; and synthesis and findings); and reaching an overall judgement of low, unclear, or high risk of bias. Each included review is appraised independently and in duplicate by two reviewers, with disagreements resolved by discussion and, where needed, adjudication by a third reviewer.

ROBIS is selected because it is a validated review-level instrument designed for systematic reviews of any type, including reviews of diagnostic accuracy, and explicitly names overviews of reviews among its target uses. Its low/unclear/high judgement maps directly onto the methodological dimension of this review's outputs, so a single appraisal output informs both the appraisal and the synthesis.

Risk of bias does not affect inclusion: no review is excluded on the basis of its ROBIS rating. Quality is instead incorporated into the synthesis, as a dimension of the review-level map, as a risk-of-bias encoding on the tool-level maps, and as a sensitivity analysis restricting the synthesis to low-risk-of-bias reviews (Item 24). Overlap between reviews at the primary-study level is quantified separately using the Corrected Covered Area method (8).

A single review-level tool is used deliberately, as there is currently insufficient empirical evidence to prefer one review-appraisal tool over another (9). AMSTAR 2 and QUADAS-2 are not used as the primary tools: AMSTAR 2 is intended for reviews of healthcare interventions (10), and QUADAS-2 appraises primary diagnostic-accuracy studies rather than reviews. Where included reviews report their own QUADAS-2 or GRADE assessments, these are extracted and presented in the supplementary tables as context. Formal GRADE/GRADE-DTA is not applied, as the umbrella-review design and the heterogeneity of tests, populations, reference standards, and thresholds across the included reviews preclude a single meaningful certainty rating. Methodological confidence is instead conveyed through the ROBIS risk-of-bias ratings, reflected in the synthesis and in the low-risk-of-bias sensitivity analysis (Item 24), with each review's own certainty statements reported narratively.

Reporting of this umbrella review follows PRIOR (11).

Strategy of data synthesis A narrative synthesis is planned, structured around the cognitive impairment continuum and organised into two complementary units of analysis: the review (to characterise the methodological evidence landscape) and the tool (to support clinical and research choices of instrument). The synthesis is anchored by evidence maps spanning the continuum stages defined in Item 10 (the four Alzheimer-spectrum stages plus the non-AD dementia parallel population column).

We intend to visualise the psychometric performance of the included tools across the continuum in separate maps for reliability, convergent validity, and diagnostic accuracy (sensitivity/specificity), complemented by a review-level map summarising the methodological coverage and risk of bias of the underlying review evidence. One plausible design plots a psychometric coefficient (for example, the ICC) on the vertical axis and the continuum on the horizontal axis, with tools shown as individual points and the strength of the supporting evidence encoded visually (for example, through point opacity). The final form of each map — including the axes, the unit summarised per cell, and any benchmarks used to aid interpretation — will be determined after data extraction, once the number of contributing tools and the sparsity of the evidence are known, and any such decisions will be reported transparently.

For each tool, the reported psychometric properties are summarised qualitatively per continuum stage, drawing on both the coefficient-based evidence and the qualitative or narrative judgements from reviews that do not enter the tool-level maps. Discrepant estimates across reviews are reported together with their range and the ROBIS ratings of the source reviews. The narrative synthesis is structured stage by stage, each stage concluding with a summary of which tools have the strongest psychometric evidence and where the gaps lie.

Given the heterogeneity of reporting formats and reference standards across the included reviews, the tool-level maps are anticipated to be sparsely populated in places, particularly for diagnostic accuracy and convergent validity and in the SCD and non-AD dementia columns. Empty cells are interpreted as gaps in the review evidence rather than as evidence of poor tool quality, and this is made explicit in the figure captions and discussion. The feasibility of the proposed maps is confirmed during a pre-extraction pilot on 2–3 anchor reviews; where a property is too sparsely populated to support a useful figure, it may be presented as a supplementary table, reported transparently.

No quantitative meta-analysis is planned. The heterogeneity of tools, stages, reference standards, and reporting formats across reviews, combined with the umbrella-review design, makes statistical pooling neither feasible nor appropriate. Heterogeneity is therefore described narratively rather than quantified (no I^2 or τ^2), and overlap among reviews at the primary-study level is quantified via the Corrected Covered Area method (Item 21).

Subgroup analysis The primary stage- and tool-level stratification (the continuum stages and the per-tool rows of the maps) is the pre-specified structuring design of the synthesis, not a subgroup analysis. Beyond it, no confirmatory subgroup analyses are pre-specified. The following are exploratory and hypothesis-generating, undertaken only where the number of supporting reviews permits and reported transparently as post hoc: delivery mode (unsupervised home vs remotely supervised), device platform (web/computer, tablet, smartphone, telephone, VR/other). Review type and ROBIS rating are not treated as subgroups here, as both are addressed as sensitivity analyses (Item 24) and, for ROBIS, as a synthesis dimension.

Sensitivity analysis Three pre-specified sensitivity analyses will test the robustness of the synthesis:
1. Restriction to ROBIS low-risk-of-bias reviews. The maps and the narrative synthesis are re-generated using only reviews rated as low risk of bias on ROBIS. Where the picture in any map changes meaningfully, this is reported and discussed.
2. Restriction to systematic reviews and meta-analyses. The tool-level maps are re-generated, excluding scoping, narrative, and literature reviews (which, on average, use less rigorous methodology than systematic reviews and meta-analyses).
3. Restriction to reviews where digital-specific psychometric data are clearly separable from non-digital data. This excludes mixed-review coefficients that required reconstruction or inference about which tools the reported psychometric data pertained to.
Each sensitivity analysis is run independently. Where the picture in the evidence maps changes meaningfully under any sensitivity analysis, this is discussed in the results and discussion sections.

Language restriction English only. Reviews not published in English will be excluded.

Country(ies) involved Netherlands.

Keywords Umbrella review; digital cognitive assessment; psychometric properties; cognitive impairment continuum; memory; mild cognitive impairment; remote neuropsychological assessment; diagnostic accuracy.

Dissemination plans The results will be submitted for publication in a peer-reviewed journal in the field of cognitive aging or neuropsychology (candidate journals include Neuropsychology Review, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, and Ageing Research Reviews), and presented at one or more national or international conferences in cognitive aging, neuropsychology, or digital health. Available as supplementary materials to the published manuscript: the PRISMA 2020 flow diagram, ROBIS ratings for each included review, the full search strings, the two-level extraction dataset (review-level and tool-level), the tool-identity harmonisation table, the count of reviews entering each map, the rationale for any review contributing to the review-level map but not to the tool-level maps, and the evidence maps (the review-level methodological-coverage map and the tool-level reliability, diagnostic accuracy, and convergent validity maps) at full resolution.

Contributions of each author

Author 1 - Joshua Panesar - CRediT: Conceptualisation; Methodology; Formal analysis; Investigation; Data curation; Writing – original draft; Writing – review & editing; Visualisation; Project administration; Software; Validation.
Email: j.r.panesar@student.rug.nl
Author 2 - Thomas Wilschut - CRediT: Conceptualisation; Methodology; Investigation; Writing – review & editing; Supervision; Validation.
Email: t.j.wilschut@rug.nl
Author 3 - Hedderik van Rijn - CRediT: Conceptualisation; Methodology; Writing – review & editing; Supervision; Resources; Funding acquisition.
Email: d.h.van.rijn@rug.nl

References 1. Jack CR, Andrews JS, Beach TG, Buracchio T, Dunn B, Graf A, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. Alzheimers Dement. 2024 Aug;20(8):5143–69. doi:10.1002/alz.13859
2. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155–63. doi:10.1016/j.jcm.2016.02.012
3. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome

measures. Qual Life Res. 2018 May;27(5):1147–57. doi:10.1007/s11136-018-1798-3
4. Terwee CB, Bot SDM, De Boer MR, Van Der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34–42. doi:10.1016/j.jclinepi.2006.03.012
5. Consensus Report of the Working Group on: "Molecular and Biochemical Markers of Alzheimer's Disease." Neurobiol Aging. 1998 Mar;19(2):109–16. doi:10.1016/S0197-4580(98)00022-0
6. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. J Thorac Oncol. 2010 Sep;5(9):1315–6. doi:10.1097/JTO.0b013e3181ec173d
7. Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016 Jan;69:225–34. doi:10.1016/j.jclinepi.2015.06.005
8. Pieper D, Antoine SL, Mathes T, Neugebauer EAM, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol. 2014 Apr;67(4):368–75. doi:10.1016/j.jclinepi.2013.11.007
9. Higgins JPT, Cochrane Collaboration, editors. Cochrane handbook for systematic reviews of interventions. Second edition. Hoboken, NJ: Wiley-Blackwell; 2019. 694 p. (Cochrane book series).
10. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017 Sep 21;j4008. doi:10.1136/bmj.j4008
11. Gates M, Gates A, Pieper D, Fernandes RM, Tricco AC, Moher D, et al. Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. BMJ. 2022 Aug 9;378:e070849. doi:10.1136/bmj-2022-070849