



Cochrane
Library

Cochrane Database of Systematic Reviews

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

Naeem F, McCleery J, Hietamies TM, Abakar Ismail F, Clinton S, O'Mahony A, Ponce OJ, Quinn TJ

Naeem F, McCleery J, Hietamies TM, Abakar Ismail F, Clinton S, O'Mahony A, Ponce OJ, Quinn TJ.
Diagnostic test accuracy of self-administered cognitive assessment tools for dementia.
Cochrane Database of Systematic Reviews 2024, Issue 12. Art. No.: CD013725.
DOI: [10.1002/14651858.CD013725.pub2](https://doi.org/10.1002/14651858.CD013725.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	8
OBJECTIVES	9
METHODS	9
Figure 1.	12
Figure 2.	13
Figure 3.	14
Test 1. TYM < 42/50	15
Test 2. TYM < 39/50	15
Test 3. TYM < 36/50	15
Test 4. SAGE	15
Test 5. CCT	15
Test 6. BrainCheck Memory	15
Test 7. p-KDSQ-C	15
Test 8. p-KDSQ-C-I	16
Test 9. p-KDSQ-C-II	16
Test 10. p-KDSQ-C-III	16
Test 11. p-KDSQ-C-IV	16
Test 12. myMemCheck	16
Test 13. TYM < 30/50	16
Test 14. TYM < 31/50	16
Test 15. TYM < 34/50	16
Test 16. TYM < 35/50	17
Test 17. TYM < 37/50	17
Test 18. TYM < 38/50	17
RESULTS	17
Figure 4.	19
Figure 5.	20
DISCUSSION	21
AUTHORS' CONCLUSIONS	23
ACKNOWLEDGEMENTS	23
REFERENCES	24
CHARACTERISTICS OF STUDIES	27
DATA	47
APPENDICES	47
HISTORY	60
CONTRIBUTIONS OF AUTHORS	60
DECLARATIONS OF INTEREST	60
SOURCES OF SUPPORT	61
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	61
INDEX TERMS	61

[Diagnostic Test Accuracy Review]

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia

Fariha Naeem¹, Jenny McCleery², Tuuli M Hietamies¹, Fatene Abakar Ismail¹, Stephanie Clinton¹, Aoife O'Mahony³, Oscar J Ponce⁴, Terry J Quinn¹

¹School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, UK. ²Oxford Health NHS Foundation Trust, Banbury, UK. ³CUBRIC, School of Psychology, Cardiff University, Cardiff, UK. ⁴Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, Minnesota, USA

Contact: Terry J Quinn, Terry.Quinn@glasgow.ac.uk.

Editorial group: Cochrane Central Editorial Service.

Publication status and date: New, published in Issue 12, 2024.

Citation: Naeem F, McCleery J, Hietamies TM, Abakar Ismail F, Clinton S, O'Mahony A, Ponce OJ, Quinn TJ. Diagnostic test accuracy of self-administered cognitive assessment tools for dementia. *Cochrane Database of Systematic Reviews* 2024, Issue 12. Art. No.: CD013725. DOI: [10.1002/14651858.CD013725.pub2](https://doi.org/10.1002/14651858.CD013725.pub2).

Copyright © 2024 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Dementia is a chronic and progressive clinical syndrome that can present with a range of cognitive and behavioural symptoms. Global prevalence is projected to increase due to ageing populations, particularly in resource-limited settings, with significant associated health and social care costs. There is a critical need for accurate cognitive assessment as part of the diagnostic workup for dementia. Although self-administered cognitive assessment tools are not diagnostic, they can be used to assess cognition. The role of these tests is uncertain, and their diagnostic test accuracy remains unclear, but they may be useful tools in circumstances where face-to-face assessment may be difficult.

Objectives

Primary objective

To assess the test accuracy of any self-administered cognitive assessment tool for the diagnosis of any form of dementia in any setting, including community and secondary health care.

Secondary objectives

To identify: the quality and quantity of the research evidence describing test accuracy of self-administered testing; sources of heterogeneity in the test accuracy described; and gaps in the evidence where further research is required.

Search methods

We searched MEDLINE (Ovid SP), Embase (Ovid), Web of Science - BIOSIS Citation Index, PsycINFO (Ovid), LILACS (BIREME), and CINAHL (EBSCO). The most recent searches were run on 2 November 2022.

Selection criteria

We included cross-sectional studies investigating the accuracy of a self-administered cognitive assessment tool. We included all settings, including community and secondary health care. The target condition of interest was a clinical diagnosis of dementia, therefore, we included only studies where the index test was administered alongside a reference standard clinical assessment. Our population of interest was any adult (over 18 years of age).

Data collection and analysis

Two review authors independently reviewed included studies, assessed risk of bias using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, and extracted data. We obtained information on study design and participant characteristics, setting of index test, details of index test, reference standard used, and results reported as sensitivity and specificity. We performed a meta-analysis on three studies that used the same threshold score.

Main results

The review included 11 eligible studies, with a total of 2303 participants, which evaluated the diagnostic test accuracy of six different self-administered cognitive-assessment tools. The studies were conducted in Europe, North America, and South Korea within a variety of community and healthcare settings. Our quality assessment found that four studies had a low risk of bias across all domains. Six studies had a high or unclear risk of bias due to patient selection, with concerns around lack of a clear sampling strategy or exclusion criteria, or both. Six studies had a high or unclear risk of bias with regard to the index test due to lack of information about how the test was conducted and evaluated.

For the diagnosis of dementia, sensitivity of self-administered cognitive assessment tools ranged from 55% to 100% and specificity ranged from 45% to 100%. Three studies described the diagnostic test accuracy of Test Your Memory at a threshold of 42/50. Quantitative meta-analysis estimated a summary point with 94% sensitivity (95% confidence interval (CI) 90% to 96%) and 66% specificity (95% CI 45% to 82%) at this threshold.

Authors' conclusions

There is insufficient evidence to recommend the use of any single self-administered cognitive assessment tool. The tools had test accuracy scores that are similar to the range seen with standard pencil and paper cognitive screening tests conducted by clinicians. Further research on the optimal test and threshold score, and how that may be impacted by setting, language, and educational level is needed.

PLAIN LANGUAGE SUMMARY

How accurate are cognitive assessment tools, that are completed by people themselves, at detecting dementia?

Key messages

- The evidence suggests that cognitive assessment tools that are completed by people themselves could be used in the detection and diagnosis of dementia.
- There is not enough evidence to recommend one type of assessment tool over another.
- Further research on the use of these assessment tools in different settings, such as clinics or people's homes, and the scores that indicate dementia is needed.

Why is improving the detection of dementia important?

The number of people diagnosed with dementia is expected to increase significantly in the future because global populations are living longer. This will have significant healthcare and social costs. Dementia is a permanent condition where memory deteriorates until people need assistance with their day-to-day activities. Accurate and quick diagnosis allows people with dementia and their families to access treatment and support. Failure to recognise the condition when it is present (a false-negative test result) may lead to a delay in accessing treatments to help with memory, thinking, and behaviours, and accessing support from social services, such as social work and occupational therapy. An incorrect diagnosis of dementia (a false-positive test) may be emotionally and psychologically distressing for people and their families. Currently, diagnosis of dementia involves a face-to-face assessment at a specialised clinic, where the medical history, a physical examination, blood tests, and brain scans are considered.

What is a self-administered cognitive assessment tool?

Self-administered cognitive assessment tools are tests completed by people themselves. They are designed to assess different aspects of mental capacity (i.e. cognition) including memory, language, and perception. The tools may involve questions to answer or tasks to complete. These self-administered tools would not replace the detailed clinical assessment required for a diagnosis of dementia, but they could be useful in situations where face-to-face assessment might be difficult, such as during global pandemics, or in areas with few medical services. There are a number of different cognitive assessment tools; these vary in their content, scoring systems, and format (electronic or written).

What did we want to find out?

We wanted to find out how accurate cognitive assessment tools that can be completed by people themselves are at detecting dementia.

What did we do?

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

Copyright © 2024 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

We searched for studies that investigated the accuracy of self-administered cognitive assessment tools for dementia in adults (over the age of 18 years) in any healthcare setting.

What did we find?

We found 11 studies with a total of 2303 participants that tested six different tools. Five studies looked at one test called 'Test Your Memory'; this was tested in five different languages. Two studies looked at the 'Self-Administered Gerocognitive Examination'; one study tested a written version of the test and the other tested an electronic version. One study looked at each of the following cognitive assessment tools individually: the 'Clock Completion Test', 'Korean Dementia Screening Questionnaire-Cognition', 'BrainCheck Memory', and 'MyMemCheck'. These tools all have different scoring systems and some studies looked at multiple cut-off scores - to indicate whether or not the person had dementia - within each tool.

Studies took place across Europe, the USA, and South Korea. The healthcare settings included community centres, nursing homes, and hospital clinics, and some studies included multiple settings. The average age of the people taking the tests varied amongst studies from 50 to 60 years to 80 to 90 years.

We found that these cognitive assessment tools could correctly identify people with dementia between 55% and 100% of the time, and could correctly rule out dementia between 45% and 100% of the time. These ranges are due to differences in study populations, settings, type of assessment tool, and diagnostic cut-off scores (i.e. score at which dementia was diagnosed). We summarised results from three studies that used 'Test Your Memory' and found that using a cut-off score of 42/50, the assessment tool could correctly identify people with dementia around 94% of the time and correctly rule out dementia around 66% of the time.

How reliable are the results of this review?

In the studies we found, the tools' diagnosis of whether a person had dementia was confirmed by a traditional, face-to-face clinical assessment for dementia. There were some problems with the methods used in some studies, including the way participants were selected for the studies, and the timings and details of the clinical assessments were not always clear. Some studies had small numbers of participants.

The test accuracy results are a summary based on evidence from the studies we examined. The results produced from individual studies varied considerably, so we cannot be sure these cognitive assessment tools will always produce the same results in clinical practice.

How up to date is this review?

The evidence is up to date to November 2022.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings: diagnostic test accuracy of self-administered cognitive assessment tools for dementia

Patient population	Any adult > 18 years of age						
Setting	Any setting including community, primary, or secondary health care						
Index test(s)	Any self-administered cognitive assessment tool completed independently by the patient						
Role and purpose of test(s)	Role uncertain, may be used in any situation where face-to-face assessment is not possible or practical within the clinical pathway of dementia assessment						
Reference standard	Any clinical diagnosis of dementia						
Target condition	Dementia (all-cause and subtypes). MCI was not one of the target conditions and was excluded as much as possible.						
Included studies	11 cross-sectional studies with 2303 eligible participants						
Quality concerns	Most studies were at high or unclear risk of bias for 1 or more domains. Six studies were at high or unclear risk of bias for patient selection due to lack of information on sampling strategy or inappropriate exclusion criteria, or both. Eight studies were at high or unclear risk of bias with regard to applicability. Six studies were at high or unclear risk of bias for the index tests with concerns around the conduct or interpretation of the index test.						
Heterogeneity	There was significant heterogeneity between studies in terms of patient population, study setting, conduct of the index tests, test thresholds, and reference standard.						
Test	Test threshold	Summary accuracy (95% CI)	Number of participants (studies)	Number of participants with target condition	Comparison	Practical implications	Quality and comments
TYM English, Norwegian and Spanish versions	42/50	Sensitivity 94% (95% CI 90% to 96%) Specificity 66% (95% CI 45% to 82%)	1289 (3)	395	Dementia versus no dementia, cognitive impairment versus no cognitive impairment	With a mean prevalence of 42% across the 3 studies, if 1000 people were tested then: - 391 people with dementia would test positive	All 3 studies recruited participants referred to outpatient clinics for cognitive assessment. However, 1 study had a low number of participants (only 16). Three different versions of the TYM were evaluated with language, cultural and contextual differences between them. MCI was treated differently.

- 29 people with dementia would test negative

- 377 people without dementia would test negative

- 203 people without dementia would test positive.

ferently across these studies. Given the heterogeneity between studies and that only 3 were included in meta-analysis, these summary estimates should be interpreted with caution.

TYM-G (Greek version)	31/50	Sensitivity 81% (95% CI 67% to 92%) Specificity 71% (95% CI 61% to 80%)	134 (1)	43	Alzheimer's disease versus no dementia	32% prevalence of AD in this study population (outpatient neurology clinics and regional hospitals). Difficult to compare to other secondary care settings due to lack of data on usual prevalence of AD.	This study had a high risk of bias surrounding the index test and its applicability due to lack of information surrounding the timing and conduct of the index test and lack of clarity in study design. MCI was not specifically excluded. There was a lack of information regarding other subtypes of dementia.
TYM-F (French version)	40/50	Sensitivity 93% (95% CI 84% to 98%) Specificity 76% (95% CI 61% to 87%)	201 (1)	68	Dementia versus no dementia (MCI excluded)	34% prevalence of dementia in this study population (outpatient memory clinic). Difficult to compare to other secondary care settings due to lack of data on usual prevalence of dementia.	Recruited new patient referrals to outpatient memory clinic. Some concerns regarding the applicability of the reference standard due to inclusion of MCI but data were reported separately with MCI excluded for certain test thresholds (not all test thresholds). Overall low risk of bias and representative population.
SAGE	16/21	Sensitivity 95% (95% CI 76% to 100%) Specificity 95% (95% CI 76% to 100%)	63 (1)	21	Dementia versus no dementia (MCI excluded)	Patients recruited from variety of community and healthcare settings. This sampling is appropriate for research but not representative for clinical services.	Overall low risk of bias across most domains. Concerns about patient selection and applicability as participants recruited from a variety of settings including community and health care with lack of information regarding the extent of cognitive concern in them. MCI excluded.
e-SAGE	14/21	Sensitivity 95% (95% CI 76% to 100%) Specificity 100% (95% CI 84% to 100%)	66 (1)	21	Dementia versus no dementia (MCI excluded)	Patients recruited from variety of community and healthcare settings. This sampling is appropriate for research but	Eligible pool of participants only included those who had previously completed SAGE. MCI excluded but no 95% CI reported. Overall low risk of bias across most domains except concerns



						not representative for clinical services.	about participant selection and applicability as per previous SAGE study.
CCT	Not specified	Sensitivity 67% Specificity 69%	53 (1)	Not specified	Alzheimer's disease versus no dementia	Patient recruited from a community setting. Number with target condition not disclosed.	Concerns about applicability of findings as study recruited only female participants, living within certain geographical areas. Lack of information about the conduct of the index test and the timing between the index test and the reference standard. 95% CI not estimable due to lack of study data.
KDSQ	Score < 3	Sensitivity 73% (95% CI 66% to 79%) Specificity 71% (95% CI 64% to 77%)	420 (1)	170	Dementia versus no dementia (MCI excluded)	40% prevalence of dementia in this study population (outpatient neurology and psychiatry clinics). Difficult to compare to other secondary care settings due to lack of data on usual prevalence of dementia.	Results displayed for p-KDSQ-C-II. Some concerns relating to patient selection and applicability of results but overall low risk of bias. MCI was excluded from results.
Braincheck Memory	Not specified	Sensitivity 89% (95% CI 52% to 100%) Specificity 78% (95% CI 40% to 97%)	18 (1)	9	Cognitive impairment versus no cognitive impairment	Over 500 participants recruited from a community setting, but only a small sub-set underwent clinical evaluation.	Small number of participants with skewed gender distribution (79.9% female). Electronic test requiring computer equipment for completion. At high or unclear risk of bias across all domains. Lack of information regarding participant selection, index test and the reference standard. Unclear how MCI was dealt with.
MyMem-Check	Not specified	Not specified	59 (1)	23	Dementia versus no dementia	Patients recruited from a community setting for an exploratory analysis as part of sub-study.	Participants recruited had either cognitive or mood concerns. High risk of bias with regard to participant selection and applicability due to lack of information on inclusion criteria. High prevalence of dementia (39%). Those with dementia scored lower (mean score = 3.91, SD = 1.79) and those without dementia scored higher (mean score = 5.25, SD = 1.90, P = 0.01). 95% CI not estimable due to lack of study data.

AD: Alzheimer's disease; CI: confidence intervals; MCI: mild cognitive impairment; SD: standard deviation

Index tests

CCT: Clock Completion Test; e-SAGE: electronic-Self-Administered Gerocognitive Exam; KDSQ: Korean Dementia Screening Questionnaire; p-KDSQ-C-II: patient-rated Korean Dementia Screening Questionnaire - Cognition items 1–5 & 11–15, memory domain + activities of daily living; SAGE: Self-Administered Gerocognitive Exam; TYM: Test Your Memory; TYM-N: Test Your Memory-Norwegian; TYM-S: Test Your Memory-Spanish; TYM-G: Test Your Memory-Greek; TYM-F: Test Your Memory-French

Note: [Appendix 1](#) provides a full overview of test results.

BACKGROUND

Dementia is a syndrome that represents the clinical manifestation of various neurodegenerative diseases, systemic conditions, and external factors that may result in brain damage. It is a chronic, progressive condition that typically affects older adults and is characterised by cognitive impairment and the inability to function independently. In 2017, 50 million people worldwide were estimated to be living with dementia; since prevalence is expected to double every 20 years, this may mean more than 130 million people may be living with dementia by 2050 ([Alzheimer's Disease International 2020](#)). As the number of sufferers has grown, so has the political profile of dementia. This is due to pressure from older people and their families, as well as the immense economic impact of dementia, which in 2015 was around 1% of global gross domestic product (GDP), mainly in costs of informal and formal social care ([Alzheimer's Disease International 2020](#)).

Assessment of cognition is a key component of dementia care. Cognitive testing can serve many important purposes ([Lin 2013](#)). In clinical practice, cognitive testing may form part of the assessment of the person with a suspected cognitive syndrome, or testing may be used as an initial screen to identify those who need more specialist input. In people with established dementia, testing may be used to assess the severity or progression of disease. In research, cognitive testing may be used to identify potential participants for a study or as an assessment of a treatment effect in a clinical trial.

Many cognitive assessment tools have been developed for different purposes and for use in a variety of settings. Indeed, in some areas there are almost as many assessment tools as there are research studies ([Lees 2012](#)). Although there is no consensus on the optimal cognitive assessment, certain tests have greater visibility and traction in research and practice. An important factor to consider when choosing a cognitive test is its accuracy for the detection of the condition of interest, for example, the accuracy of a screening test for detection of dementia.

In the Cochrane Dementia and Cognitive Improvement Group (CDCIG), we have reviewed the literature and summarised the accuracy of many of the commonly used cognitive screening tests, including Folstein's Mini-Mental State Examination ([Creavin 2016](#)), Montreal Cognitive Assessment ([Davis 2015](#)), and Addenbrooke's Cognitive Examination ([Beishon 2019](#)). Our diagnostic test accuracy (DTA) reviews have largely been focused on in-person, face-to-face assessment, as this would be usual practice in most services ([Davis 2013](#)). However, there are also cognitive assessment tools that have been designed for self-administration without the involvement of a clinician. Here it is important to state that whilst we talk of 'diagnostic' test accuracy, and use clinical dementia diagnosis as a comparator, the cognitive assessment tools of interest in this review are not sufficient to make a dementia diagnosis on their own.

These cognitive assessment tools have been used mostly in research and their role in clinical practice is uncertain. They could potentially be used in situations where face-to-face assessment is difficult. The coronavirus pandemic caused a fundamental change in practice that no one had anticipated. The emergency restrictions on movement and social contact necessitated by the viral pandemic limited the opportunity for in-person assessment. This has focused interest on assessment tools which do not require face-to-face contact with a clinician - often termed 'remote

assessment'. Beyond the pandemic situation, there are other circumstances in which self-administered cognitive tests could be useful clinically, for example, in rural areas or in resource-poor services ([Barth 2018](#)), and, in the research context, such tests may offer a feasible way to include cognitive outcome measures in large pragmatic trials or observational studies ([Ritchie 2015](#)).

A variety of cognitive screening tests designed for self-administration is available ([Charalambous 2020](#); [Koo 2019](#)). However, the diagnostic accuracy of these tools should not be assumed, so we felt it necessary to collate, appraise, and present estimates of accuracy for studies describing the use of self-administered cognitive assessment tools. Companion reviews to this review which use similar methods examine the accuracy of:

- cognitive tests which are administered remotely by a clinician, using telephone or video technology ([Beishon 2022](#)); and
- a remote telemedicine-based dementia assessment service ([McCleery 2021](#)).

We hope that this suite of reviews will help inform remote assessment of people with cognitive concerns.

Target condition being diagnosed

The condition of interest for this review was clinical dementia. We recognise that cognitive testing may be used to inform the diagnosis of other cognitive syndromes such as mild cognitive impairment (MCI), but the condition of greatest relevance is invariably clinical dementia.

The dementia diagnosis is operationalised in various classification systems, such as the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) (although in the latest iteration of the DSM, the term 'major neurocognitive disorder' rather than 'dementia' is used) and the World Health Organization's International Classification of Diseases (ICD) ([APA 2013](#); [WHO 2010](#)). Although there are differences between these classifications, they all describe dementia as a clinical syndrome characterised by impairments in multiple cognitive domains sufficient to cause problems in activities of daily living. Additional classifications exist to describe pathological dementia subtypes, for example Alzheimer's disease or vascular dementia.

Index test(s)

Our index test of interest was any cognitive assessment tool, usually questionnaire style - but may include cognitive tasks or tests - that is completed independently by the patient, that is, without the supervision of a healthcare professional, and which is intended to screen for dementia or form part of an assessment of suspected dementia. Assessment tools administered with a small degree of supervision by a layperson (e.g. a relative) were eligible. Eligible index tests include those designed to be completed on paper or on a computer, tablet, or other mobile device. The purpose of the index test should be cognitive assessment and in our index test conditions we did not include 'serious gaming' or cognitive assessment that is derived from performance on another test or game.

Clinical pathway

For consistency with other reviews, and in keeping with usual Cochrane methods, we use the phrase 'diagnostic test accuracy'.

However, we recognise that our index tests of interest are not sufficient to make a diagnosis on their own. In clinical practice a cognitive assessment tool is often the first step in a diagnostic process. This is followed by a detailed, multidisciplinary assessment that may also be informed by assessments of function, collateral history, and radiological and laboratory testing (Noel-Storr 2012). The clinical assessment of dementia often takes place within secondary healthcare settings, namely psychiatry, neurology and geriatric medicine clinics.

Role of index test(s)

The role of these self-administered cognitive assessment tools in clinical practice is uncertain. Our interest, for the purpose of this review, includes evaluating the use of these tools in any setting. These self-administered tools could potentially be used in a variety of situations. This may include use in community settings for a variety of screening purposes. People concerned about possible dementia may complete publicly available cognitive tests, including commercial tests, to help them to decide whether to consult a healthcare professional (e.g. a primary care physician). Self-administered cognitive assessment tools may be used as a screening tool for identifying people in the community who may be suitable for inclusion in dementia research studies. They could be used for public health purposes, for example, distributed to everyone within a particular geographical area and demographic category to aid planning of dementia services. They could also be used to screen for dementia in particular clinical populations that may struggle to attend for in-person health care, for example post-COVID-19. It is less likely that self-administered cognitive assessment tools would be used as part of the clinical assessment of suspected dementia in primary or secondary care. However, such use may occur particularly in any situation where access to face-to-face clinical assessment is difficult (e.g. during a pandemic, in remote and rural areas, or in resource-poor settings where few trained staff are available).

Alternative test(s)

An alternative to self-administered cognitive tests is testing administered remotely by a clinician using telephone or video technology. These approaches are the subject of a separate review (Beishon 2022). There are many short cognitive screens that involve in-person assessment and Cochrane have reviewed the accuracy of some of the most commonly used tests (Beishon 2019; Creavin 2016; Davis 2015). There are also questionnaires that assess for change in cognition and function that can be completed by a collateral source, for example, family or caregiver (Harrison 2015; Hendry 2019).

Rationale

The arguments for timely diagnosis of dementia have been made by various professional societies (Robinson 2015). Suffice it to say, cognitive testing is fundamental to the assessment of the person with a suspected cognitive problem. Our motivation for this review was in response to the recent global viral pandemic, where best practice of in-person cognitive testing was often not safe or possible. Clinical and research teams had to adapt rapidly to new ways of working, and we hope to provide an evidence base to guide choice of testing. We also believe that the push for greater use of remote testing may continue despite COVID-19 restrictions on movement being lifted.

In keeping with other Cochrane reviews, the property of interest in this work is the accuracy of the test. We recognise that accuracy is only one of several factors that should be considered when applying a test strategy; equally important are factors such as the implications of the test result, the feasibility of testing (Lees 2017), or the effect of tests on subsequent clinical pathways (Takwoingi 2018). There may be potential to explore these in future reviews, but they would require differing approaches to the classical test accuracy paradigm.

OBJECTIVES

To assess the test accuracy of any self-administered cognitive assessment tool for the diagnosis of any form of dementia in any setting, including community and secondary health care.

Secondary objectives

To identify:

- the quality and quantity of research evidence describing test accuracy of self-administered testing;
- sources of heterogeneity in the test accuracy described; and
- gaps in the evidence where further research is required.

METHODS

Criteria for considering studies for this review

Types of studies

We included cross-sectional studies where the index test was administered alongside reference standard clinical assessment.

We included studies that took place in any setting including the community and healthcare settings, due to the uncertainty of the role of these index tests in clinical practice.

We excluded case-control studies due to their inherent risk of bias and because these data cannot be used to make any inferences about population predictive value. For the same reasons, we excluded studies that used an enriched sample, for example, that only included participants who already had a certain screening test score, or where the reference standard assessment was limited to participants with a particular cognitive profile.

Studies where the index test is compared against the future development of a cognitive syndrome (delayed-verification studies) require a different review approach compared to the traditional cross-sectional test accuracy study. We therefore did not consider delayed-verification studies in this review.

We did not include studies where the index and reference standard were intended to be administered with more than one month between them. Such studies should be considered prognostic, given that the prognostic test accuracy of the index test in predicting future development of disease is being analysed. Studies where the planned delay between index and reference was one month or less were potentially eligible and the potential effect of this delay was considered as part of the risk of bias assessment.

We excluded studies with a small number of cases (fewer than 10), as these studies are unlikely to add meaningfully to our understanding of test accuracy and are prone to a variety of selection biases.

Participants

Our population of interest was any adult (age over 18 years).

We did not include studies that were composed exclusively of cognitively 'normal' participants, which are used to create normative values of tests.

We did not exclude studies on the basis of a selected population, but, where the population was not predominantly an older adult group (e.g. studies in traumatic brain injury or studies in stroke), we noted this and, where possible, explored case-mix as a source of heterogeneity.

Index tests

Studies included - but not necessarily exclusively - a self-administered cognitive assessment tool that was completed independently by the participant without supervision from a healthcare professional. A small degree of supervision from a layperson (e.g. a relative) was acceptable. These tools included written and electronic formats.

We anticipated including several index tests to perform indirect comparisons of estimates of accuracy (Owen, 2018), but there was an insufficient number of studies to support such analyses.

Target conditions

We considered studies that reported any clinical diagnosis of dementia. Dementia diagnosis could be undifferentiated, or a particular subtype could be specified. Although classifying dementia by subtype was not required for inclusion, where available, these data were recorded. Some studies reported data on mild cognitive impairment (MCI) as well as clinical dementia and, where possible, the data on MCI were excluded.

We did not include studies that reported only on MCI (Langa 2014), or delirium (Hendry 2016), in this review, as these cognitive syndromes are less suitable for the brief self-administered testing that was the focus of this review.

Reference standards

Our reference standard is a clinical diagnosis of dementia. Within the clinical diagnosis rubric, we included all-cause (unspecified) dementia, made using any recognised diagnostic criteria, for example ICD-10 or DSM-IV. For the purposes of this review, and in keeping with other Cochrane diagnostic test accuracy (DTA) reviews, we considered structured interview assessments such as Clinical Dementia Rating (CDR) as diagnostic (Davis 2013). Dementia diagnoses may specify a pathological subtype, and all dementia subtypes were included.

Studies that made a post-mortem diagnosis or diagnosis based on imaging or other biomarkers without a corresponding comprehensive clinical assessment were not eligible. We did not set any limits in relation to severity or stage of dementia. We had planned to explore severity of dementia as a potential source of heterogeneity, but these data were not available.

Search methods for identification of studies

Electronic searches

We searched:

- MEDLINE Ovid (1950 to 2 November 2022);
- Embase Ovid (1974 to 2 November 2022);
- CINAHL EBSCO (Cumulative Index to Nursing and Allied Health Literature; 1982 to 2 November 2022);
- PsycINFO Ovid (from 1806 to 2 November 2022);
- LILACS (Latin American and Caribbean Health Science Information database; 1982 to 2 November 2022); and
- Web of Science - BIOSIS Citation Index (1950 to 2 November 2022).

Each database was searched from inception year to the date of the search. These databases allow for a search of the grey literature that is appropriate to the topic of this review. See [Appendix 2](#) for strategies run. We used controlled vocabulary such as MeSH terms and Emtree where appropriate. In the searches developed, we made no attempt to restrict studies on the basis of sampling frame or setting. This approach was intended to maximise sensitivity and allow inclusion on the basis of population-based sampling to be assessed at screening (see [Selection of studies](#)). We did not use search filters (collections of terms aimed at reducing the number needed to screen) as an overall limiter because those that are published have not proved sensitive enough (Whiting 2008). We did not apply any language restriction to the electronic searches, and used translation services as necessary. We searched ALOIS, the Cochrane Dementia and Cognitive Improvement Group's (CDCIG) Specialised Register, which includes both intervention and DTA studies in dementia. When we found potentially relevant studies, we performed citation analysis, including forward and backward citation searching.

Searching other resources

We did not search any other resources.

Data collection and analysis

Selection of studies

Following searching, titles from various databases were collated in Covidence software (Covidence 2020). A single review author performed a 'first pass' review, removing clearly irrelevant titles, then a minimum of two review authors independently assessed studies for eligibility.

For consistency with our other DTA titles, we adopted a hierarchical approach to exclusion, first excluding on the basis of index test and reference standard, then on the basis of study methods (case-control, size), and then for any other reason.

Where a potentially relevant study did not include the data needed for analyses, we attempted to contact the primary author by email twice. If the authors did not respond, or the relevant data were not available, we could not include the study in any quantitative synthesis. If the same dataset was presented in more than one study, we included the primary publication and made reference to other publications if they contained relevant additional information.

Where studies were described in abstract form, we contacted the lead author(s) to ask if the full study was published. We limited the studies included to those published in peer-reviewed scientific journals.

We detailed the study selection process in a PRISMA flow diagram ([Figure 1](#)).

Figure 1. PRISMA flow diagram.

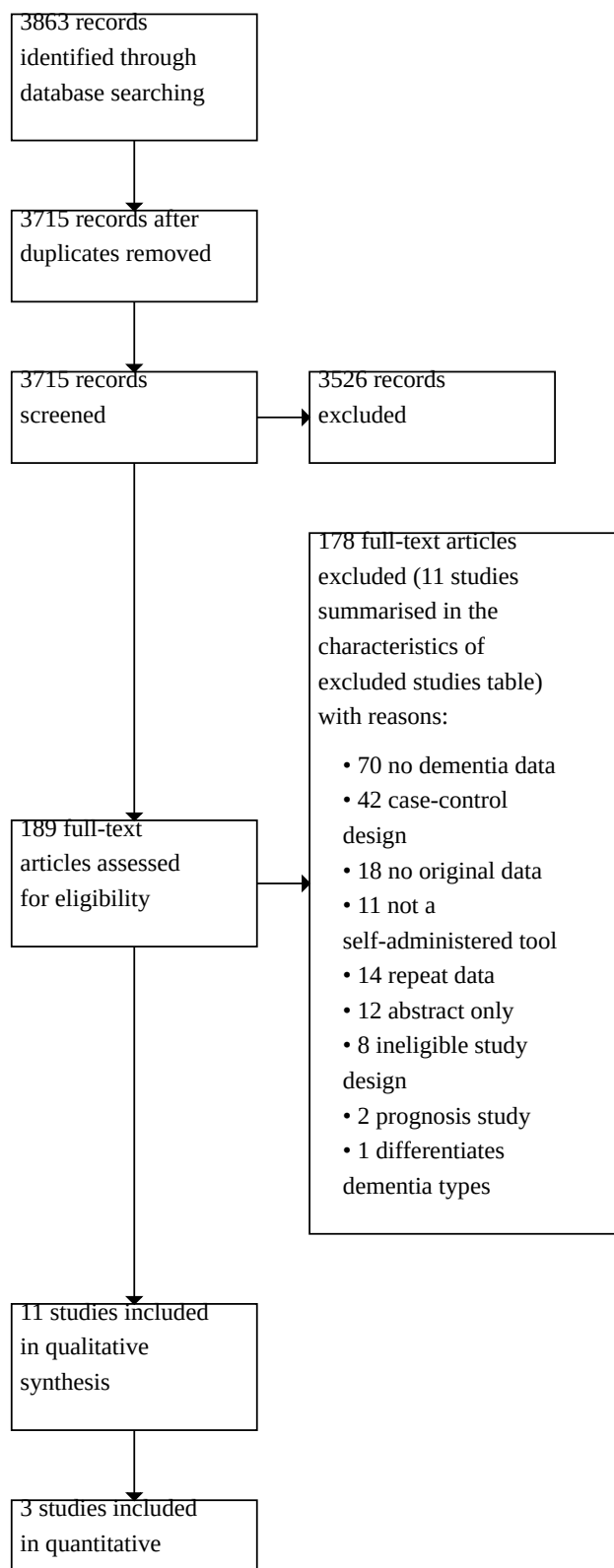


Figure 1. (Continued)

in quantitative
synthesis
(meta-analysis)

Data extraction and management

Two review authors independently extracted data from the eligible studies onto a bespoke data extraction form. The form described the population tested, purpose/setting of testing, test(s) administered, details of reference standard assessment and order of testing (cognitive assessment tool first or clinical assessment first). We collected the data needed to create a 2 x 2 table including prevalence of dementia. The utility of questionnaire-based assessments can be limited by poor return rates so we also collected data on completion and return rates. Many of the included studies collected data on MCI as an outcome as well as dementia. Where possible, data were collected on dementia versus no dementia with MCI excluded as a subgroup. If this information was not available, then data relating to dementia versus no dementia was collected. In some studies the only information available concerned outcomes relating to cognitive impairment versus no cognitive impairment. This information was summarised in [Appendix 1](#).

We piloted the data extraction form on two studies and made any required changes. Following data extraction, the two review authors compared findings and resolved any disagreements through discussion, with recourse to a senior test accuracy review author as needed.

Where a test assigned a score and accuracy data were given for a variety of threshold scores, we collected all these data in the

first instance. Primary analysis was limited to performance at the standard threshold for that test (where a standard exists). Exploratory analyses described accuracy at other thresholds. We created a summary of findings table to summarise the results of the included studies ([Summary of findings 1](#)).

Assessment of methodological quality

Paired, independent raters blinded to each other's scores assessed risk of bias (internal validity) and generalisability (external validity) using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool ([Whiting 2011](#)).

QUADAS-2 assessment covers issues that relate to patient selection, index test, reference standard, and participant flow. Each domain is assessed for issues with risk of bias; the first three domains are also assessed for concerns about generalisability. Our group has considerable experience using the QUADAS-2 tool. For previous DTA reviews, we convened a group with expertise in test accuracy and dementia to tailor the QUADAS-2 approach to the field of dementia screening-test studies. Through this work we have operationalised scoring rules for item- and domain-level QUADAS-2 assessment, and modified the generic QUADAS-2 question stems to cover issues pertinent to cognitive testing (details of the process informing our QUADAS assessment are available in [Appendix 3](#)) ([Davis 2013](#)). We presented our QUADAS-2 results as graphical displays ([Figure 2](#); [Figure 3](#)), and as a narrative within the text.

Figure 2. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.

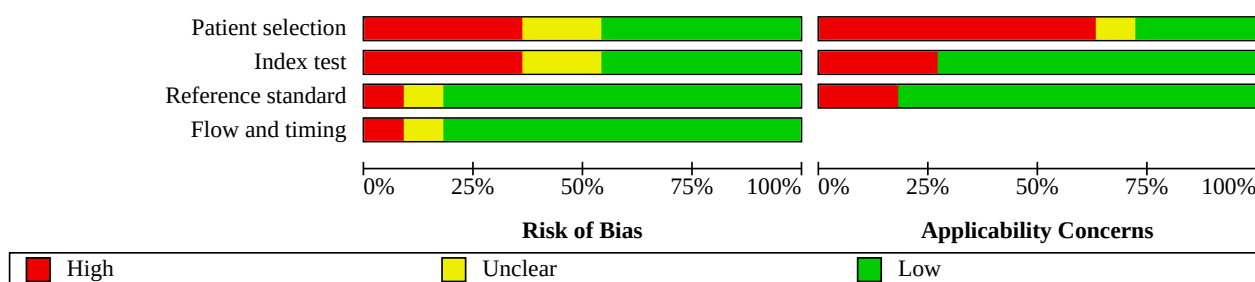


Figure 3. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

	Risk of Bias				Applicability Concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ball 2001	?	—	+	—	—	+	+
Breitve 2015	?	?	?	+	?	—	+
Ferrero-Arias 2016	—	—	+	+	—	+	+
Groppell 2019	—	—	—	?	—	—	—
Hancock 2010	+	+	+	+	+	+	+
Iatraki 2014	+	—	+	+	+	—	+
Kim 2019	—	?	+	+	—	+	+
Mansbach 2020	—	+	+	+	—	+	+
Postel-Vinay 2014	+	+	+	+	+	+	—
Scharre 2010	+	+	+	+	—	+	+
Scharre 2017	+	+	+	+	—	+	+

High
 Unclear
 Low

Statistical analysis and data synthesis

Our primary analysis of interest was accuracy of the various self-administered assessments against the dichotomous outcome variable 'dementia/no dementia'. To explore this, we applied the

recommended Cochrane framework treating each test separately. For each test, where data allowed, we extracted data to populate a standard 2 x 2 data table of binary test results (above and below threshold score) cross-classified against binary reference standard.

From this table we calculated sensitivities and specificities, with 95% confidence intervals, at individual test level for thresholds of interest. Where there was no common threshold, we performed this process for all clinically plausible thresholds or thresholds where data were available. We presented these study results graphically in a forest plot of sensitivity and corresponding specificity ([Data table 1](#);

[Data table 2](#); [Data table 3](#); [Data table 4](#); [Data table 5](#); [Data table 6](#); [Data table 7](#); [Data table 8](#); [Data table 9](#); [Data table 10](#); [Data table 11](#); [Data table 12](#); [Data table 13](#); [Data table 14](#); [Data table 15](#); [Data table 16](#); [Data table 17](#); [Data table 18](#)). We performed these first analyses with standard Review Manager software ([RevMan 2024](#)).

Test 1. TYM < 42/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Breitve 2015	10	1	0	5	1.00 [0.69, 1.00]	0.83 [0.36, 1.00]		
Hancock 2010	74	80	4	66	0.95 [0.87, 0.99]	0.45 [0.37, 0.54]		

Test 2. TYM < 39/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Postel-Vinay 2014	61	40	7	93	0.90 [0.80, 0.96]	0.70 [0.61, 0.78]		

Test 3. TYM < 36/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ferrero-Arias 2016	119	101	8	821	0.94 [0.88, 0.97]	0.89 [0.87, 0.91]		

Test 4. SAGE

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Scharre 2010	20	1	1	20	0.95 [0.76, 1.00]	0.95 [0.76, 1.00]		
Scharre 2017	20	0	1	21	0.95 [0.76, 1.00]	1.00 [0.84, 1.00]		

Test 5. CCT

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ball 2001	0	0	0	0	Not estimable	Not estimable		

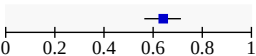
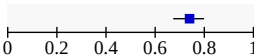
Test 6. BrainCheck Memory

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Groppell 2019	8	2	1	7	0.89 [0.52, 1.00]	0.78 [0.40, 0.97]		

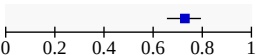
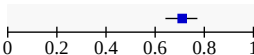
Test 7. p-KDSQ-C

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kim 2019	105	46	65	154	0.62 [0.54, 0.69]	0.77 [0.71, 0.83]		

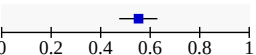
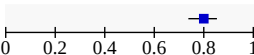
Test 8. p-KDSQ-C-I

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kim 2019	109	52	61	148	0.64 [0.56, 0.71]	0.74 [0.67, 0.80]		

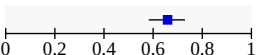
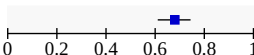
Test 9. p-KDSQ-C-II

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kim 2019	124	58	46	142	0.73 [0.66, 0.79]	0.71 [0.64, 0.77]		

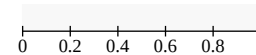
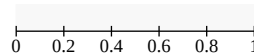
Test 10. p-KDSQ-C-III

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kim 2019	94	40	76	160	0.55 [0.47, 0.63]	0.80 [0.74, 0.85]		

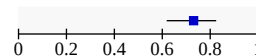
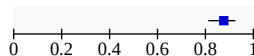
Test 11. p-KDSQ-C-IV

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kim 2019	112	64	58	136	0.66 [0.58, 0.73]	0.68 [0.61, 0.74]		

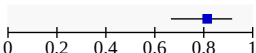
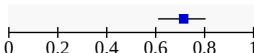
Test 12. myMemCheck

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Mansbach 2020	0	0	0	0	Not estimable	Not estimable		

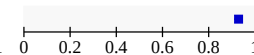
Test 13. TYM < 30/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Hancock 2010	57	18	21	128	0.73 [0.62, 0.82]	0.88 [0.81, 0.93]		

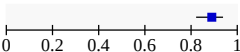
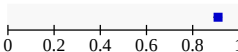
Test 14. TYM < 31/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Iatraki 2014	35	26	8	65	0.81 [0.67, 0.92]	0.71 [0.61, 0.80]		

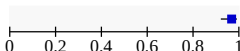
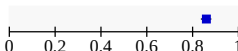
Test 15. TYM < 34/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ferrero-Arias 2016	104	65	23	857	0.82 [0.74, 0.88]	0.93 [0.91, 0.95]		

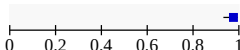
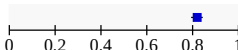
Test 16. TYM < 35/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ferrero-Arias 2016	113	83	14	839	0.89 [0.82, 0.94]	0.91 [0.89, 0.93]		

Test 17. TYM < 37/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ferrero-Arias 2016	123	129	4	793	0.97 [0.92, 0.99]	0.86 [0.84, 0.88]		

Test 18. TYM < 38/50

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ferrero-Arias 2016	124	166	3	756	0.98 [0.93, 1.00]	0.82 [0.79, 0.84]		

Where there were more than two studies describing a test with the same threshold value, we attempted quantitative meta-analysis using random-effects models. In the first instance we created summary estimates of sensitivity and specificity using the bivariate approach. For this we used bespoke software MeaDTA (Patel 2020).

Investigations of heterogeneity

We did not quantify statistical heterogeneity for these DTA analyses but presented 95% prediction intervals around any summary estimate and the ranges of sensitivity and specificity that informed the estimate.

We reviewed forest plots of sensitivity and specificity looking for outliers.

Heterogeneity is often seen in clinical test accuracy reviews (Deeks 2001). We planned various subgroup analyses to assess the potential effect of common sources of heterogeneity in dementia test accuracy. Planned analyses included populations tested, setting of testing, technical aspects of the testing (e.g. language of the assessment tool), and criteria used to make the dementia diagnosis. The limited number of returned studies precluded formal, quantitative analysis, but we describe these aspects for the various tests included.

Sensitivity analyses

We also planned to explore the effect of methodological aspects of the included studies, running a sensitivity analysis limited to those studies at low risk of bias. Due to the few numbers of included studies, our exploration of bias was through a narrative description.

Assessment of reporting bias

We chose not to perform a quantitative assessment of reporting bias. We recognise that debate remains around the most robust approach to assessment of reporting bias in DTA (Wilson 2015), and there is uncertainty about how to apply standard approaches such as funnel plots (van Enst 2014).

RESULTS

Results of the search

The search resulted in 3715 citations after initial de-duplication. Following title and abstract screening, 189 eligible studies underwent full-text review; 178 of these articles were excluded (11 of which are summarised in the [Characteristics of excluded studies](#) table). Reasons for exclusion were: existing as an abstract only, case-control study, no dementia data, not a self-administered tool, no original data, or data not suitable for analysis (see [Figure 1](#)).

This review includes 11 studies ([Characteristics of included studies](#)) with a total of 2303 eligible participants (range: 16 to 1049). Some studies conducted multiple trials and only those participants who underwent a self-administered screening test and were assessed against a reference standard were included.

Methodological quality of included studies

We used the QUADAS-2 tool to assess the 11 included studies for methodological quality and risk of bias. One study was rated at low risk of bias for all categories of QUADAS-2 (Hancock 2010). QUADAS-2 results have been presented as graphical displays and as a narrative within the text, see [Figure 2](#); [Figure 3](#).

Patient selection and sampling

We assessed five studies as being at low risk of bias (Hancock 2010; Iatraki 2014; Postel-Vinay 2014; Scharre 2010; Scharre 2017), and two studies at unclear risk of bias (Ball 2001; Breivte 2015). The remaining four studies we judged to be at high risk of bias due to lack of information about sampling strategy and/or exclusion criteria or inappropriate exclusion criteria (Ferrero-Arias 2016; Groppe 2019; Kim 2019; Mansbach 2020).

Five studies recruited participants referred to outpatient clinics (Ferrero-Arias 2016; Hancock 2010; Iatraki 2014; Kim 2019; Postel-Vinay 2014). Three studies were community based including nursing home residents (Ball 2001; Groppe 2019; Mansbach 2020),

and two studies were a combination of participants recruited from clinic and community (Scharre 2010; Scharre 2017).

In terms of applicability, we assessed three studies as being at low risk of bias (Hancock 2010; Iatraki 2014; Postel-Vinay 2014), as they recruited patients referred to outpatient clinics due to cognitive concerns. The remaining studies were either unclear or at high risk of bias with regard to applicability, as they recruited people from the community, or outpatient clinics, without specifying whether there were any pre-existing concerns about cognitive decline.

Index test application

Four studies had a high risk of bias with regard to the conduct or interpretation of the index test (Ball 2001; Ferrero-Arias 2016; Groppell 2019; Iatraki 2014). Two of these studies had no specified thresholds for their cognitive assessment tools (Ball 2001; Groppell 2019), in one study the cognitive assessment tool was scored after the reference standard assessment by the same individual (Ferrero-Arias 2016), and in one study there was a lack of clarity surrounding the timing and conduct of the index test (Iatraki 2014). We assessed two studies as having an unclear risk of bias (Breitve 2015; Kim 2019). The rest of the studies were at low risk of bias. Two studies mentioned a threshold score of 42/50 for Test Your Memory (TYM) (Hancock 2010; Iatraki 2014), but the rest of the studies had no prespecified threshold.

For three studies there was high concern about the applicability of the index test (Breitve 2015; Groppell 2019; Iatraki 2014); we assessed the rest of the studies as being of low concern.

Reference standard (clinical diagnosis of dementia) application

One study was perceived to be at high risk of bias due to a lack of information regarding the components of the reference standard test (Groppell 2019), and another was assessed as having an unclear risk of bias (Breitve 2015). In two studies we had high concern about applicability, because the reference standard was used to diagnose mild cognitive impairment rather than dementia (Groppell 2019; Postel-Vinay 2014).

Flow and timing

We assessed one study as being at high risk of bias for this domain (Ball 2001), as there was a delay of at least two weeks between the participants completing the index test and being invited for a clinical assessment. One study was deemed to have an unclear risk of bias (Groppell 2019), and the rest of the studies we considered to be at low risk of bias.

Findings

We identified 11 international studies that compared the accuracy of a self-administered cognitive assessment to a clinical dementia diagnosis. The details of these studies are summarised in [Characteristics of included studies](#) and the findings displayed in the [Summary of findings 1](#). One study was reported in Norwegian (Breitve 2015), and all other studies were reported in English.

The self-administered cognitive assessment tools utilised in the 11 studies were: Test Your Memory (TYM), which was used in five different languages based on the native language of the study population, Self-Administered Gerocognitive Exam (SAGE), Korean Dementia Screening Questionnaire - Cognition (KDSQ-

C), BrainCheck Memory, MyMemCheck, and an isolated two-page questionnaire devised by the authors of the Ball 2001 study that included a Clock Completion Test (CCT).

For the diagnosis of dementia, sensitivity of self-administered cognitive assessment tools ranged from 55% to 100% and specificity ranged from 45% to 100% (see [Appendix 1](#) for full list of results).

Test Your Memory (TYM)

The TYM was designed in 2009 to be available for widespread use by a non-specialist with three critical requirements in mind: a short administration time, testing a range of cognitive functions, and being sensitive to mild Alzheimer's disease (Brown 2009). It is a written assessment comprised of 11 tasks including orientation, ability to copy a sentence, semantic knowledge, calculation, verbal fluency, similarities, naming, visuospatial abilities, and recall of a copied sentence. The eleventh task is the ability to complete the test. This gives a potential maximum of 50 points.

Five studies in this review used TYM translated into their native language (Breitve 2015; Ferrero-Arias 2016; Hancock 2010; Iatraki 2014; Postel-Vinay 2014).

Target condition

In three studies all-cause dementia and mild cognitive impairment (MCI) were the target conditions (Ferrero-Arias 2016; Hancock 2010; Postel-Vinay 2014), one study looked at Alzheimer's dementia (AD) and MCI (Iatraki 2014), and one study investigated dementia and AD as a specific subtype (Breitve 2015).

Setting

All five studies were conducted in secondary care and involved participants referred to outpatient hospital settings, including neurology and old age psychiatry clinics.

Threshold

Only one study had a predefined test threshold of 42/50 (Breitve 2015). At this threshold the Norwegian version of the TYM test had a sensitivity of 100% and a specificity of 56% for diagnostic test accuracy of detecting AD and a sensitivity of 100% and a specificity of 83% for detecting all cases of dementia. Another study found that at a threshold of 42 the sensitivity of TYM for detecting dementia was 95%, but with a suboptimal specificity of 45% (Hancock 2010). At a threshold of 30, the sensitivity was 73% and specificity was 88%. A Spanish version of TYM was evaluated at multiple test thresholds (Ferrero-Arias 2016). At a threshold of 36, the test had a sensitivity of 94% and specificity of 89% for detecting dementia. This score was corrected according to the educational level of the participants. The team demonstrated that as the test threshold increased, the sensitivity of the test increased but the specificity decreased. A French version of TYM was investigated and the team found a score of 40 had a 93% sensitivity and 76% specificity in diagnostic test accuracy for detecting dementia versus no dementia (MCI excluded) (Postel-Vinay 2014). Lastly, one study devised a Greek version of TYM and used age- and education-adjusted test thresholds (Iatraki 2014). At a threshold of 31, the test had a sensitivity of 81% and specificity of 71% for detecting AD.

Quantitative meta-analysis

Three studies described the diagnostic test accuracy of TYM at the same threshold value of 42/50 (Breitve 2015; Ferrero-Arias 2016; Hancock 2010). We performed a quantitative meta-analysis to estimate a summary point at this common threshold using random-effects models and the bespoke software MetaDTA (Patel

2020), as displayed in Figure 4. Three datasets contained relevant data ($n = 1289$). The sensitivity was 94% (95% confidence interval (CI) 90% to 96%); specificity was 66% (95% CI 45% to 82%). The prevalence of dementia in this particular population was 42%. Using this prevalence, we have presented a visual representation of test accuracy at this test threshold (Figure 5).

Figure 4. Summary estimate for TYM data at threshold of 42/50 Key Circle = data point Diamond = summary estimate Greyed out area = 95% confidence region Dotted line = 95% predictive region

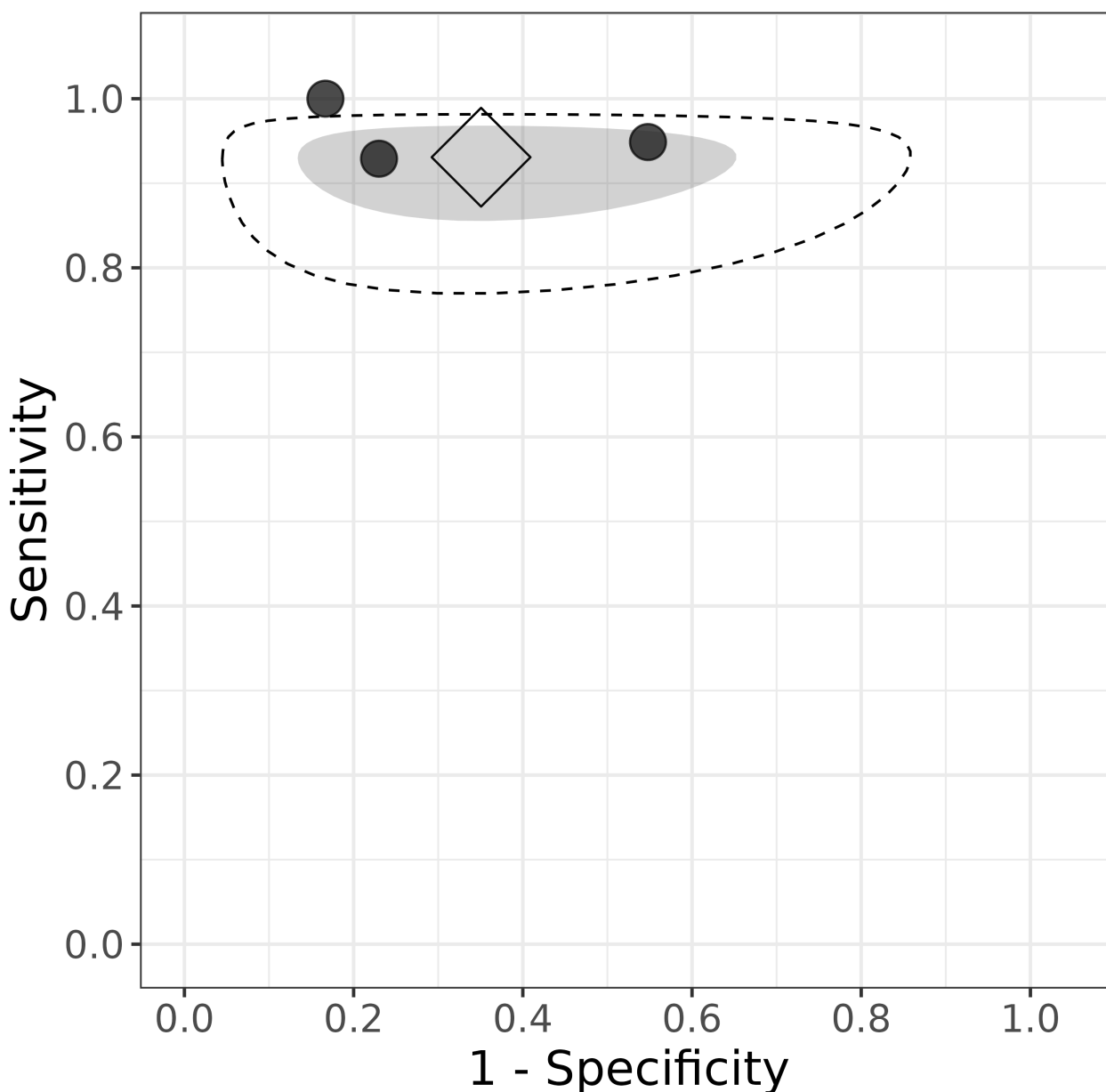
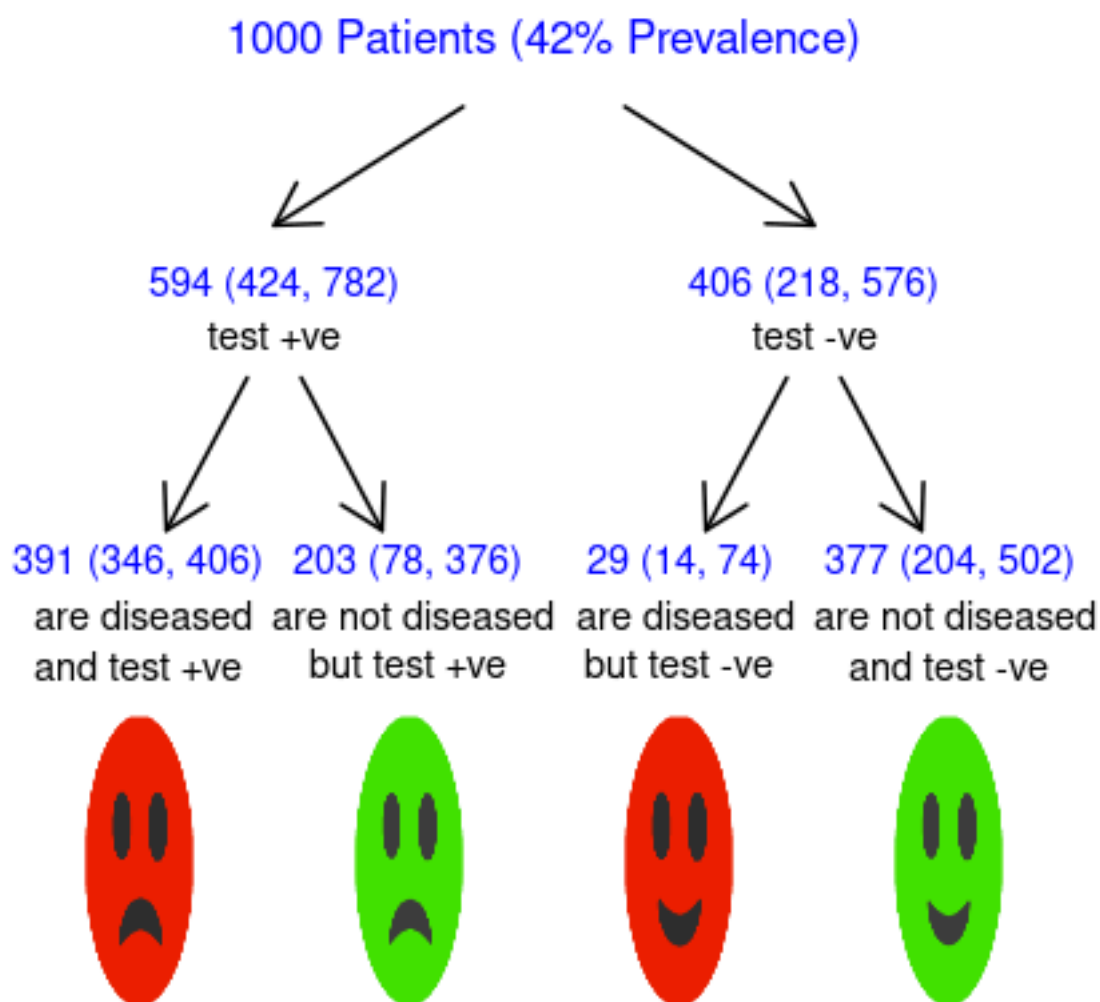


Figure 5. Prevalence of dementia in the population (as a percentage) Note: the numbers in brackets represent 95% confidence intervals.



Self-Administered Gerocognitive Examination (SAGE)

SAGE is a written 22-point self-assessment, that consists of questions in seven domains of cognitive functioning: orientation, language, memory, executive function, calculations, abstraction, and visuospatial abilities. There are also 'non-scoring items' that are intended to assist in clinician identification of cognitive impairment. These include participant's educational level achieved, past medical history of stroke, family history of cognitive impairment, and any current symptoms of cognitive decline.

One study used a written version ([Scharre 2010](#)), and a further study by the same research team looked at an electronic version of the same questionnaire (eSAGE) ([Scharre 2017](#)).

Target condition

Target conditions were dementia and MCI for both studies.

Setting

Participants were recruited from a wide variety of settings including geriatric and memory disorder outpatient clinics, senior centres, educational talks and newspaper adverts.

Threshold

For the written version of SAGE a score of 16 provided a sensitivity of 95% and specificity of 95% in detecting dementia versus no dementia (MCI excluded). For eSAGE a score of 14 gave a 95% sensitivity and 100% specificity for the diagnostic test accuracy of detecting dementia versus no dementia (MCI excluded).

Clock Completion Test (CCT)

One study used the 'clock drawing test' as the screening method for Alzheimer's-type dementia (AD) (Ball 2001). This is a written screening assessment, completed by patients at home, as part of a two-page questionnaire that includes general information about health history and questions to assess Alzheimer's disease risk factors. The CCT was scored as either being normal or abnormal.

Target condition

The target condition was AD.

Setting

Community-based females aged 65 years or older living in prespecified geographical areas.

Threshold

An abnormal CCT score had a 67% sensitivity and 69% specificity for detection of AD. This was improved by the addition of two or more risk factors for AD to 75% sensitivity and 87% specificity.

Korean Dementia Screening Questionnaire-Cognition (KDSQ-C)

The KDSQ-C is a 15-item semi-structured questionnaire designed to be completed by reliable informants or caregivers.

One study constructed a shortened form of the KDSQ-C for self-assessment purposes and tested its reliability and validity for use as a screening tool for cognitive impairment (Kim 2019).

Target condition

Cognitive impairment including dementia and MCI.

Setting

Patients were recruited from neurology or psychiatry outpatient clinics as well as regional dementia centres.

Threshold

Data were collected for different subsets of the KDSQ-C and whether the test was completed by the participant (p) or caregiver (i). In terms of best discriminability between normal cognition and dementia groups the p-KDSQ-C-II had the best sensitivity and specificity at 73% and 71%, respectively, with a test threshold of 3.

BrainCheck Memory

BrainCheck Memory is a modified version of BrainCheck Sport; a computerised neurocognitive test developed to detect concussion (Yang 2017). BrainCheck Memory functions as an app which can be downloaded and accessed by a password-protected login. It tests seven neurocognitive domains: immediate and delayed recall, the Trail Making Test A, the Trail Making Test B, the Stroop Test, the Digit Symbol Substitution Task, and matrix problems task.

One study aimed to validate BrainCheck Memory (Groppell 2019).

Target condition

The target condition was age-related cognitive impairment.

Setting

This study was community based with participants recruited from assisted living facilities, community centres, and a church.

Threshold

Compared to the reference standard of physician diagnosis, BrainCheck Memory had a sensitivity of 89% and specificity of 78% in detecting moderate to severe cognitive impairment.

MyMemCheck

This screening tool was established as a rapid cognitive self-assessment tool to improve early diagnosis of mild cognitive impairment and dementia. It includes items on both functional and cognitive performance, to increase patient participation and identify early cognitive changes. Eight items are scored either 'correct' or 'incorrect', with a total maximum score of 8. It is completed on paper.

One study investigated the use of this short written questionnaire (Mansbach 2020).

Target condition

Dementia and MCI.

Setting

Participants were community dwelling residents from independent living facilities and nursing homes.

Threshold

The study found that participants with a diagnosis of dementia, as per the reference standard, scored significantly lower on the screening test ($n = 23$, mean score = 3.91, standard deviation (SD) = 1.79) than those without dementia ($n = 36$, mean score = 5.25, SD = 1.90, $P = 0.01$). Data were not presented to allow us to extrapolate the specificity or sensitivity of the screening test.

DISCUSSION

Summary of main results

This review identified 11 studies that looked at the diagnostic test accuracy of six different screening tools for detecting dementia. Common reasons for study exclusion were ineligible study design and lack of data about dementia diagnosis. Most of these screening tools came as written assessments with two screening tools that were to be completed electronically. There was significant heterogeneity between the study populations in terms of mean age, proportion of female participants, and educational level (when it was specified). Six studies had relatively small sample sizes of less than 100 participants, four studies had sample sizes between 100 and 500, and only one study had a sample size of more than 1000 participants (Ferrero-Arias 2016).

We deemed four studies as having a low risk of bias across all domains (Hancock 2010; Postel-Vinay 2014; Scharre 2010; Scharre 2017). The highest risk of bias amongst studies was noted to

be in patient selection, which reflects the heterogeneity within the populations from which study participants were recruited. Some studies recruited patients referred to outpatient clinics, some recruited from community centres and residential areas, and others recruited from community and healthcare settings. We noted that some studies had inappropriate or unclear exclusion criteria (Ball 2001; Breitve 2015; Ferrero-Arias 2016; Groppell 2019; Kim 2019; Mansbach 2020). Similarly, we had concerns around applicability, as these participants may not accurately reflect real life populations. Some studies we noted had a high risk of bias relating to the index tests (Ball 2001; Ferrero-Arias 2016; Groppell 2019; Iatraki 2014). Two studies did not specify the scoring threshold for their index tests (Ball 2001; Groppell 2019), while another two studies had a lack of information regarding the scoring and timing of the index test in comparison to the reference standard (Ferrero-Arias 2016; Iatraki 2014), which raised concerns that these factors might have affected the interpretation of the index tests. This may have implications for the test accuracy data for these tests, which included the Clock Completion Test (CCT), BrainCheck, and Test Your Memory (TYM) (Spanish and Greek versions), as well as the reproducibility of their results.

Sensitivity for the diagnosis of dementia through self-administered cognitive assessment ranged from 55% to 100%, and specificity ranged from 45% to 100%. There were differences in target condition across the studies, with inclusion of mild cognitive impairment (MCI) in some analyses, lack of consensus on optimum test thresholds, and heterogeneity within study populations and study settings, which may account for some of this variability.

These results are comparable to those from other Cochrane diagnostic test accuracy reviews in this field.

- One evaluated the Mini-Mental State Examination (MMSE), which is a widely used in-person screening test for dementia, and found it had a sensitivity of 85% (95% confidence interval (CI) 74% to 92%) and specificity of 90% (95% CI 82% to 95%) at a threshold of 24/30 (Creavin 2016).
- Another found that the Addenbrooke's Cognitive Examination III (ACE-III) had a sensitivity of between 75% to 97% depending on the test threshold and setting, with even greater variability in specificity (50% to 77% in outpatient memory clinic) (Beishon 2019). With regard to remote assessment by video or telephone, this review found that sensitivity of these tests varied between 26% and 100% and specificity between 65% and 100% (Beishon 2022). The setting, education level of participants, and test threshold were all factors that contributed to this variability. No studies were found that directly compared the performance of self-administered cognitive assessment tools and standard screening tests in the same cohort.

Strengths and weaknesses of the review

The strengths of this review include that we conducted it in line with a prespecified protocol. This is in accordance with guidance published on undertaking a diagnostic test accuracy review of cognitive assessment tools (Davis 2013). The comprehensive search strategy was developed by Information Specialists at Cochrane. Each stage of the review was carried out independently by two assessors. If data were missing or unclear we contacted the study authors for clarification.

The weaknesses of this review include the small number of studies we identified, and the significant heterogeneity between the studies with regard to the study populations and the way the index tests were conducted, which may limit the interpretation of these findings. There was a lack of prespecified test thresholds. The reference standard also varied amongst studies with an assortment of tools being used to diagnose dementia including blood tests, neuroimaging, and clinical assessment. This introduces a degree of subjectivity to the assessment of cognitive impairment and functional decline. There was a lack of information on dementia subtypes and severity of dementia. The role of MCI and its inclusion in results and analysis also varied widely amongst the studies.

Applicability of findings to the review question

The primary objective of this review was to assess the test accuracy of any self-administered cognitive assessment tool for the diagnosis of any form of dementia. All included studies focused on a written or electronic cognitive screening tool that was completed by the participant. This was then compared with a clinical diagnosis of dementia.

The populations in the studies varied considerably. Mean age ranged from 50 to 60 years to 80 to 90 years. Research was also conducted across a variety of geographical areas encompassing different languages and cultures. Some studies took place in community, religious, and residential settings, and targeted older adults who may or may not have had concerns about their cognition. Other studies took place in secondary care clinics, often neurology or memory clinics, and targeted referrals to these clinics. Referral to neurology clinics could include concerns about cognition, but could also include neurological symptoms with no cognitive involvement and some studies made reference to this in their analyses. We did not differentiate between the types of clinic as different specialties may be involved in the assessment of cognition. Sensitivity of these assessment tools was generally higher than specificity in most studies. The setting has important implications when it comes to testing and the clinical role of these cognitive assessment tools. In the community and residential settings where these self-administered cognitive assessment tools may be used to screen or triage populations for clinical purposes, a higher sensitivity could be desirable. A higher specificity may be important when considering screening a population for research participants with a particular condition. In healthcare settings such as secondary care clinics, cognitive assessment forms one aspect of a multi-domain assessment and a higher specificity may be more useful.

These results would not be applicable to special populations, for example, motor neurone disease or stroke. Many studies also collected and reported data on MCI as well as dementia. Our exclusion of those with MCI may have reduced the applicability of our findings. People with MCI are likely to be amongst those offered cognitive assessment, so exclusion could have led to overestimation of the diagnostic accuracy of these assessment tools. However, these subgroup data were not always excluded from subsequent analyses and often treated differently between studies, with some including MCI in the 'no dementia' outcome and other studies including MCI along with dementia in the 'cognitive impairment' outcome. Whilst the main outcome of interest in this review was dementia, the impact of those with MCI on the findings of this review is difficult to estimate.

Most studies used written screening tools, which is the format most likely to be adopted in clinical practice. Screening tools that use modern technology may be useful in more remote or rural settings. No study explored the cost, feasibility, or sustainability of their screening tools or any additional resources that might be needed.

AUTHORS' CONCLUSIONS

Implications for practice

Currently, there is insufficient evidence to recommend the use of any single self-administered cognitive assessment tool. Test Your Memory has been evaluated in multiple studies and has now been translated into several different languages. The extent to which educational level and regional differences in culture were taken into consideration varied amongst all the studies. There is also a lack of consensus on test threshold.

The issues of ageing populations, increased demand within healthcare services, and limited resources highlight the potential value of self-administered cognitive assessment. The tools had test accuracy scores similar to the range seen with standard pencil and paper cognitive screening tests conducted by clinicians. Studies have been conducted in both community settings, with a presumed relatively low prevalence of dementia, and secondary care settings where the prevalence would be expected to be higher. These tests have a good sensitivity, but specificity remains variable, highlighting their utility as a screening tool in addition to expert clinical history and diagnosis.

Implications for research

This review highlights the need for further research to establish the accuracy of self-administered cognitive screening tools, their optimum test thresholds, assessment characteristics and how these may vary in different settings, including community and secondary healthcare settings. The heterogeneity within participant populations (community dwelling older adults or those with cognitive concerns referred to outpatient clinics), cultural context and impact of native languages, educational level, and age also need to be explored further, as well as any adjustments that may need to be applied to scores. A direct comparison between self-administered tools and standard screening tests would also be useful. Ideally, the index test and reference test should be completed on the same day, independently of each other. Consensus on the reference standard test to reduce subjectivity would be beneficial. Further research into detection of dementia subtypes and cognitive impairment in specific circumstances, such as post stroke, is also needed.

With a focused approach on establishing the optimal setting, test threshold, and target population, the role of these self-administered cognitive assessment tools in clinical practice may become clearer.

ACKNOWLEDGEMENTS

Editorial and peer-reviewer contributions

Cochrane Dementia and Cognitive Improvement supported the authors in the development of this diagnostic test accuracy review.

The Cochrane Diagnostic Test Accuracy (DTA) editorial team and Central Editorial Service editorial team co-ordinated a dual editorial process.

The following people conducted the editorial process for this article:

- Sign-off Editor (final editorial decision): Professor Leon Flicker, Western Australian Centre for Health & Ageing – WACHA, University of Western Australia, Perth, Western Australia (clinical); Professor Sascha Köpke, Institute of Nursing Science, University of Cologne, Germany (DTA);
- Managing Editor (selected peer reviewers, provided editorial guidance to authors, edited the article): Luisa Fernandez Mauleffinch, Cochrane Central Editorial Service;
- Editorial Assistant (conducted editorial policy checks, collated peer-reviewer comments and supported editorial team): Lisa Wydrzynski, Cochrane Central Editorial Service;
- Copy Editor (copy editing and production): Elizabeth Royle, Production Manager, Cochrane Central Production Team;
- Peer-reviewers (provided comments and recommended an editorial decision):
 - Matthew J Grainge, School of Medicine, Division of Epidemiology and Public Health, University of Nottingham, UK (DTA statistical review);
 - Daniëlle AWM van der Windt, Department of Primary Care & Health Sciences, Keele University, UK (DTA methods review);
- Cochrane DTA editorial team (search review);
 - Clinical Professor Constance Dimity Pond, Wicking Dementia Research and Teaching Centre, University of Tasmania (clinical review);
 - Kasper Jørgensen, Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital-Rigshospitalet, Copenhagen, Denmark (clinical review); and
 - Catherine Hofstetter, Consumer (consumer review).

REFERENCES

References to studies included in this review

Ball 2001 {published data only} <https://doi.org/10.1046/j.1532-5415.2001.49159.x>

* Ball LJ, Ogden A, Mandi D, Birge SJ. The validation of a mailed health survey for screening of dementia of the Alzheimer's type. *Journal of the American Geriatrics Society* 2001;**49**(6):798-802. [DOI: doi.org/10.1046/j.1532-5415.2001.49159.x]

Breitve 2015 {published data only}

Breitve MH, Chwiszczuk L, Hynninen M. A Norwegian pilot study of the TYM test (Test Your Memory). *Norsk Psykologforening* 2015;**52**(1):49-53.

Ferrero-Arias 2016 {published data only}

Ferrero-Arias J, Turrión-Rojo MÁ. Validation of a Spanish version of the Test Your Memory. *Neurologia* 2016;**31**(1):33-42.

Groppell 2019 {published data only}

Groppell S, Soto-Ruiz KM, Flores B, Dawkins W, Smith I, Eagleman DM, et al. A rapid, mobile neurocognitive screening test to aid in identifying cognitive impairment and dementia (BrainCheck): cohort study. *JMIR Aging* 2019;**2**(1):e12615.

Hancock 2010 {published data only}

Hancock P, Larner AJ. Test Your Memory test: diagnostic utility in a memory clinic population. *International Journal of Geriatric Psychiatry* 2011;**25**:976-80.

Iatraki 2014 {published data only}

Iatraki E, Simos PG, Lionis C, Zaganas I, Symvoulakis EK, Papastefanakis E, et al. Cultural adaptation, standardization and clinical validity of the Test Your Memory dementia screening instrument in Greek. *Dementia and Geriatric Cognitive Disorders* 2014;**37**:163-80.

Kim 2019 {published data only}

Kim A, Kim S, Park KW, Park KH, Youn YC, Lee DW, et al. Reliability and validity of a short form of the Korean Dementia Screening Questionnaire-Cognition. *Journal of Clinical Neurology* 2020;**16**(1):143-53.

Mansbach 2020 {published data only (unpublished sought but not used)}

Mansbach WE, Mace RA, Tanner MA, Simmons SP. A path to early diagnosis of mild cognitive impairment and dementia: validity and reliability of the myMemCheck® self-administered screening tool. *Family Practice* 2020;**37**(4):561-7.

Postel-Vinay 2014 {published data only}

Postel-Vinay N, Hanon O, Clerson P, Brown JM, Ménard J, Paillaud E, et al. Validation of the Test Your Memory (F-TYM Test) in a French memory clinic population. *Clinical Neuropsychologist* 2014;**28**(6):994-1007.

Scharre 2010 {published data only}

Scharre DW, Chang SI, Murden RA, Lamb J, Beversdorf DQ, Kataki M, et al. Self-administered Gerocognitive Examination (SAGE): a brief cognitive assessment Instrument for mild

cognitive impairment (MCI) and early dementia. *Alzheimer Disease and Associated Disorders* 2010;**24**(1):64-71. [DOI: [10.1097/WAD.0b013e3181b03277](https://doi.org/10.1097/WAD.0b013e3181b03277)] [PMID: 20220323]

Scharre 2017 {published data only} [10.1186/s13195-017-0269-3](https://doi.org/10.1186/s13195-017-0269-3)

Scharre DW, Chang SI, Nagaraja HN, Vrettos NE, Bornstein RA. Digitally translated Self-Administered Gerocognitive Examination (eSAGE): relationship with its validated paper version, neuropsychological evaluations, and clinical assessments. *Alzheimer's Research and Therapy* 2017;**9**(1):44. [DOI: [10.1186/s13195-017-0269-3](https://doi.org/10.1186/s13195-017-0269-3)] [PMID: 28655351]

References to studies excluded from this review

Brown 2017 {published data only}

Brown JM, Lansdall CJ, Wiggins J, Dawson KE, Hunter K, Rowe JB, et al. The Test Your Memory for Mild Cognitive Impairment (TYM-MCI). *Journal of Neurology, Neurosurgery and Psychiatry* 2017;**88**(12):1045-51.

Crane 2012 {published data only (unpublished sought but not used)}

Crane M, Dougherty J, Cannon R. Use of the computer self-test to detect difference in cognitive patterns between AD and FTD. *Dementia and Geriatric Cognitive Disorders* 2012;**34**:235.

Dong 2019 {published data only (unpublished sought but not used)}

Dong Y, Ling MT, Ng KE, Wang A, Wan EY, Merchant RA, et al. The clinical utility of the TYM and RBANS in a one-stop memory clinic in Singapore: a pilot study. *Journal of Geriatric Psychiatry and Neurology* 2019;**32**(2):68-73.

Gallassia 2012 {published data only (unpublished sought but not used)}

Gallassia R, Guttmanb S, Monaldinib C, De Matteisa M, Gallassia V, Sambatia L. Symptoms of dementia screener: is it a useful questionnaire for cognitive impairment screening? *Journal of Alzheimer's Disease* 2012;**29**(Supp 1):67-8.

Galvin 2007 {published data only}

Galvin JE, Roe CM, Coats MA, Morris JC. Patient's rating of cognitive ability: using the AD8, a brief informant interview, as a self-rating tool to detect dementia. *Archives of Neurology* 2007;**64**:725-30.

Jorgensen 2020 {published data only}

Jorgensen K, Nielsen TR, Nielsen A, Waldorff FB, Waldemar G. Brief Assessment of Impaired Cognition Questionnaire (BASIC-Q) - development and validation of a new tool for identification of cognitive impairment in community settings. *International Journal of Geriatric Psychiatry* 2020;**35**:693-701.

Larner 2015 {published data only}

Larner AJ. Hard-TYM: a pragmatic study. *International Journal of Geriatric Psychiatry* 2015;**30**(3):330-1.

Miyamae 2016 {published data only}

Miyamae F. The development of a self-administered dementia checklist: the examination of concurrent validity and discriminant validity. *Japanese Journal of Geriatrics* 2017;**53**(4):354-62.

Saunders 2008 {published data only}

Saunders LV. Validation of a self-administered computerized screening test for cognitive impairment and dementia. *Neurology* 2008;**70**(11, Supp 1):A184-5.

Tsai 2018 {published data only}

Tsai PH, Liu JL, Lin KN, Chang CC, Pai MC, Wang WF, et al. Development and validation of a dementia screening tool for primary care in Taiwan: Brain Health Test. *PLOS One* 2018;**13**(4):e0196214.

Van Mierlo 2017 {published data only}

Van Mierlo LD. Screening for mild cognitive impairment and dementia with automated, anonymous online and telephone cognitive self-tests. *Journal of Alzheimer's Disease* 2017;**56**:249-59.

Additional references
Alzheimer's Disease International 2020

Alzheimer's Disease International. Dementia statistics. www.alz.co.uk/research/statistics (accessed prior to 8 September 2020).

APA 2013

American Psychiatric Association, DSM-5 Task Force. Diagnostic and statistical manual of mental disorders: DSM-5. 5th edition. Vol. **xliv**. Washington, DC: American Psychiatric Association, 2013.

Barth 2018

Barth J, Nickel F. Diagnosis of cognitive decline and dementia in rural areas - a scoping review. *Journal of Geriatric Psychiatry* 2018;**33**(3):459-74.

Beishon 2019

Beishon LC, Batterham AP, Quinn TJ, Nelson CP, Panerai RB, Robinson T, et al. Addenbrooke's Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia and mild cognitive impairment. *Cochrane Database of Systematic Reviews* 2019, Issue 12. Art. No: CD013282. [DOI: [10.1002/14651858.CD013282.pub2](https://doi.org/10.1002/14651858.CD013282.pub2)]

Beishon 2022

Beishon LC, Elliott E, Hietamies TM, McArdle R, O'Mahony A, Elliott AR, et al. Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia. *Cochrane Database of Systematic Reviews* 2022, Issue 4. Art. No: CD013724. [DOI: [10.1002/14651858.CD013724.pub2](https://doi.org/10.1002/14651858.CD013724.pub2)]

Brown 2009

Brown J, Pengas G, Dawson K, Brown LA, Clatworthy P. Self administered cognitive screening test (TYM) for detection of

Alzheimer's disease: cross sectional study. *BMJ* 2009;**338**:b2030. [DOI: [10.1136/bmj.b2030](https://doi.org/10.1136/bmj.b2030)]

Charalambous 2020

Charalambous AP, Pye A, Yeung WK, Leroi I, Neil M, Thodi C, et al. Tools for app- and web-based self-testing of cognitive impairment: systematic search and evaluation. *Journal of Medical Internet Research* 2020;**22**(1):e14451.

Covidence 2020 [Computer program]

Covidence. Melbourne, Australia: Veritas Health Innovation, 2020. Available at www.covidence.org.

Creavin 2016

Creavin ST, Wisniewski S, Noel-Storr AH, Trevelyan CM, Hampton T, Rayment D, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of Systematic Reviews* 2016, Issue 1. Art. No: CD011145. [DOI: [10.1002/14651858.CD011145.pub2](https://doi.org/10.1002/14651858.CD011145.pub2)]

Davis 2013

Davis DH, Creavin ST, Noel-Storr A, Quinn TJ, Smailagic N, Hyde C, et al. Neuropsychological tests for the diagnosis of Alzheimer's disease dementia and other dementias: a generic protocol for cross-sectional and delayed-verification studies. *Cochrane Database of Systematic Reviews* 2013, Issue 3. Art. No: CD010460. [DOI: [10.1002/14651858.CD010460](https://doi.org/10.1002/14651858.CD010460)]

Davis 2015

Davis DH, Creavin ST, Yip JL, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane Database of Systematic Reviews* 2015, Issue 10. Art. No: CD010775. [DOI: [10.1002/14651858.CD010775.pub2](https://doi.org/10.1002/14651858.CD010775.pub2)]

Deeks 2001

Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**:157-62.

Harrison 2015

Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. *Cochrane Database of Systematic Reviews* 2015, Issue 3. Art. No: CD123456. [DOI: [10.1002/14651858.CD123456](https://doi.org/10.1002/14651858.CD123456)]

Hendry 2016

Hendry K, Quinn TJ, Evans J, Scortichini V, Miller H, Burns J, et al. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. *Age and Ageing* 2016;**45**:832-7. [DOI: [10.1093/ageing/afw130](https://doi.org/10.1093/ageing/afw130)]

Hendry 2019

Hendry K, Green C, McShane R, Noel-Storr AH, Stott DJ, Anwer S, et al. AD-8 for detection of dementia across a variety of healthcare settings. *Cochrane Database of*

Systematic Reviews 2019, Issue 3. Art. No: CD123456. [DOI: [10.1002/14651858.CD123456](https://doi.org/10.1002/14651858.CD123456)]

Koo 2019

Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. *Innovation in Aging* 2019;**3**(1):1-14.

Langa 2014

Langa KM, Levine DA. The diagnosis and management of mild cognitive impairment. *JAMA* 2014;**312**:2551-61.

Lees 2012

Lees R, Fearon P, Harrison JK, Broomfield NK, Quinn TJ. Cognitive and mood assessment in stroke research: focused review of contemporary studies. *Stroke* 2012;**43**(6):1678-80.

Lees 2017

Lees RA, Hendry K, Broomfield N, Stott DJ, Larner AJ, Quinn TJ. Cognitive assessment in stroke: feasibility and test properties using differing approaches to scoring of incomplete items. *International Journal of Geriatric Psychiatry* 2017;**32**:1072-72.

Lin 2013

Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: a systematic review for the US Preventive Services Task Force. *Annals of Internal Medicine* 2013;**159**(9):601-12.

McCleery 2021

McCleery J, Laverty J, Quinn TJ. Diagnostic test accuracy of telehealth assessment for dementia and mild cognitive impairment. *Cochrane Database of Systematic Reviews* 2021, Issue 7. Art. No: CD013786. [DOI: [10.1002/14651858.CD013786.pub2](https://doi.org/10.1002/14651858.CD013786.pub2)] [PMID: 34282852]

Noel-Storr 2012

Noel-Storr AH, McCleery JM, Richard E, Ritchie CW, Flicker L, Cullum SJ, et al. Reporting standards for studies of diagnostic test accuracy in dementia: the STARDdem Initiative. *Neurology* 2012;**83**(4):364-73.

Patel 2020

Patel A, Cooper N, Freeman S, Sutton A. Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data. *Research Synthesis Methods* 2020;**12**(1):34-44. [DOI: doi.org/10.1002/jrsm.1439]

RevMan 2024 [Computer program]

Review Manager (RevMan). Version 7.14.1. The Cochrane Collaboration, 2024. Available at revman.cochrane.org.

Ritchie 2015

Ritchie CW, Terrera GM, Quinn TJ. Dementia trials and dementia tribulations: methodological and analytical challenges

in dementia research. *Alzheimer's Research & Therapy* 2015;**7**(1):31.

Robinson 2015

Robinson L, Tang E, Taylor JP. Dementia: timely diagnosis and early intervention. *BMJ* 2015;**350**:h3029.

Takwoingi 2018

Takwoingi Y, Quinn TJ. Review of Diagnostic Test Accuracy (DTA) studies in older people. *Age and Ageing* 2018;**47**:349-55.

van Enst 2014

van Enst AW, Ochodo E, Scholten RJ, Hooft L, Leeftang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Medical Research Methodology* 2014;**14**:70.

Whiting 2008

Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. *Journal of Clinical Epidemiology* 2008;**41**(4):357-64.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeftang MM, Sterne JA, Bossuyt PM, . QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 18 Oct 2011;**155**(8):529-36. [DOI: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)]

WHO 2010

World Health Organization. International Statistical Classification of Diseases and Related Health Problems (ICD). Vol. 2. Geneva: World Health Organization, 2010.

Wilson 2015

Wilson C, Kerr D, Noel-Storr A, Quinn TJ. Associations with publication and assessing publication bias in dementia diagnostic test accuracy studies. *International Journal of Geriatric Psychiatry* 2015;**30**(12):1250-6.

Yang 2017

Yang S, Flores B, Magal R, Harris K, Gross J, Ewbank A, et al. Diagnostic accuracy of tablet-based software for the detection of concussion. *PLOS One* 2017;**12**(7):e0179352.

References to other published versions of this review

Quinn 2020

Quinn TJ, McCleery J, Hietamies TM, Abakar Ismail F. Diagnostic test accuracy of self-administered cognitive assessment questionnaires for dementia. *Cochrane Database of Systematic Reviews* 2020, Issue 9. Art. No: CD013725. [DOI: [10.1002/14651858.CD013725](https://doi.org/10.1002/14651858.CD013725)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Ball 2001

Study characteristics			
Patient Sampling	400 women were recruited from a community-based educational programme for older adults. <u>Sampling procedure</u> : random <u>Inclusion criteria</u> : living independently within selected zip codes based on socioeconomic diversity, > 65 years old, female <u>Exclusion criteria</u> : not specified		
Patient characteristics and setting	Community setting A sample of 400 women were selected. 95 were excluded due to having died, not being contactable, or not meeting the inclusion criteria. Of the remaining 305 women, 191 returned the health survey and 53 of these respondents then went on to have a clinical evaluation. <u>Total number of participants</u> : 305 <u>Number of participants included in qualitative analysis</u> : 53 <u>Mean age (years)</u> : not specified <u>Sex (% female)</u> : 100 <u>Education (years)</u> : mean 14.4 +2.8		
Index tests	2-page questionnaire, including the Clock Completion Test, questions about general health and to assess for risk factors for AD		
Target condition and reference standard(s)	<u>Target condition</u> : AD <u>Reference standard</u> : NINCDS-ADRDA criteria for probable AD		
Flow and timing	There were at least 2 weeks between completing the index test (2-page questionnaire) and being invited to clinical assessment.		
Comparative			
Notes	Study authors contacted in 2022 to enquire about total number of patients diagnosed with dementia. No response was given.		
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Ball 2001 (Continued)

Did the study avoid inappropriate exclusions?	Unclear	
Could the selection of patients have introduced bias?		Unclear risk
Are there concerns that the included patients and setting do not match the review question?		High
DOMAIN 2: Index test (All tests)		
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes	
If a threshold was used, was it pre-specified?	No	
Could the conduct or interpretation of the index test have introduced bias?		High risk
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 3: Reference standard		
Is the reference standard likely to correctly classify the target condition?	Yes	
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Breitve 2015

Study characteristics		
Patient Sampling	16 patients who were referred for dementia assessment at the neurology and old age psychiatry departments at Haugesund Hospital were recruited.	
	Sampling procedure: not specified	
	Inclusion criteria: not specified	

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

Breitve 2015 (Continued)

	Exclusion criteria: not specified
Patient characteristics and setting	Outpatient setting Total number of participants: 16 Number of participants included in qualitative analysis: 16 Mean age (years): 66.5 (range 52-89) Sex (% female): not specified Education: not specified
Index tests	TYM-N (Norwegian version)
Target condition and reference standard(s)	Target conditions: dementia and AD Reference standard: ICD-10
Flow and timing	TYM and reference tests were performed in the same day.
Comparative	
Notes	TYM in this study was done under supervision of healthcare professional, but the degree of assistance was not specified.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference standard			

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

Breitve 2015 (Continued)

Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern
DOMAIN 4: Flow and timing	
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Ferrero-Arias 2016

Study characteristics	
Patient Sampling	1049 participants were recruited over a 3.5 year period. These were new patients referred to a particular author's neurology clinic. <u>Sampling procedure</u>: consecutive <u>Inclusion criteria</u>: age > 50 years <u>Exclusion criteria</u>: sensory or motor impairment, illiteracy, < 4 years of formal education, chronic alcoholism, major psychiatric disorders, those taking significant amounts of psychotropic substances.
Patient characteristics and setting	Outpatient setting The group was divided into 3 groups: - those with dementia - those with MCI - those considered to be 'cognitively normal' consisting of controls (carers/relatives/non-neurological referrals to clinics), patients with neurological disorders but no brain involvement and patients with neurological disorders and brain involvement but no cognitive impairment. <u>Total number of participants</u>: 1049 <u>Number of participants included in qualitative analysis</u>: 1049 <u>Mean age (years)</u>: dementia 79.2+8.0, MCI 76.7+7.2, 'cognitively normal' 71.8+9.3 <u>Sex (% female)</u>: dementia 70%, MCI 57%, 'cognitively normal' 60%

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

Ferrero-Arias 2016 (Continued)

Education (% less than 13 years education): dementia 75.6%, MCI 67.4%, 'cognitively normal' 57.4%

Index tests	TYM-S (Spanish version)
Target condition and reference standard(s)	<u>Target conditions</u> : dementia, MCI <u>Reference standards</u> : DSM-IV and CDR (clinical dementia rating) stage 1 or 2 for dementia, Petersen criteria for MCI and CDR stage 0.5.
Flow and timing	TYM test was performed first, prior to the clinical evaluation and on the same day. The same person reviewed and scored the TYM after having established the clinical diagnosis.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Ferrero-Arias 2016 (Continued)

Could the reference standard, its conduct, or its interpretation have introduced bias?

Low risk

Are there concerns that the target condition as defined by the reference standard does not match the question?

Low concern

DOMAIN 4: Flow and timing

Was there an appropriate interval between index test and reference standard?

Yes

Did all patients receive the same reference standard?

Yes

Were all patients included in the analysis?

Yes

Could the patient flow have introduced bias?

Low risk

Groppell 2019
Study characteristics

Patient Sampling

583 participants were recruited from community centres, assisted living facilities and a church in Houston, USA.

Sampling procedure: not specified

Inclusion criteria: age ≥ 50 years, function in at least 1 hand, normal or corrected vision

Exclusion criteria: history of stroke or other neurological disability, inability to speak English or Spanish, illiteracy defined as inability to read the written informed consent.

Patient characteristics and setting

Community setting

A subset of study participants (18 residents) underwent the reference standard test and were eligible for inclusion in our analyses.

Total number of participants: 583

Number of participants in qualitative analysis: 18

Mean age (years): 85.9 (SD 7.3)

Sex (% female): not specified

Education: not specified

Index tests

BrainCheck

Target condition and reference standard(s)

Target condition: age related cognitive impairment

Reference standard: not stated explicitly and explained as 'personal and medical history followed by administration of the MoCA test'

Flow and timing

BrainCheck test and clinical assessment were performed on the same day.

Groppell 2019 (Continued)

Comparative

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

33

Groppell 2019 (Continued)

Could the patient flow have introduced bias?

Unclear risk

Hancock 2010
Study characteristics

Patient Sampling	224 participants who were referred to 2 memory clinics over a 23-month period were recruited to the study. <u>Sampling procedure</u>: consecutive <u>Inclusion criteria</u>: not specified explicitly but study suggests all new referrals were eligible <u>Exclusion criteria</u>: not specified explicitly but study suggests all new referrals were eligible
Patient characteristics and setting	Outpatient setting <u>Total number of participants</u>: 224 <u>Number of participants in qualitative analysis</u>: 224 <u>Mean age (years)</u>: 63.3 ±12.6, range 20-90 <u>Sex (% female)</u>: 42% <u>Education</u>: not specified
Index tests	TYM
Target condition and reference standard(s)	<u>Target conditions</u>: dementia, MCI <u>Reference standard</u>: DSM-IV
Flow and timing	TYM and clinical assessment were performed on the same day but independently of each other.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	

Hancock 2010 (Continued)

Are there concerns that the included patients and setting do not match the review question? Low concern

DOMAIN 2: Index test (All tests)

Were the index test results interpreted without knowledge of the results of the reference standard? Yes

If a threshold was used, was it pre-specified? Yes

Could the conduct or interpretation of the index test have introduced bias? Low risk

Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern

DOMAIN 3: Reference standard

Is the reference standard likely to correctly classify the target condition? Yes

Were the reference standard results interpreted without knowledge of the results of the index tests? Yes

Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk

Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern

DOMAIN 4: Flow and timing

Was there an appropriate interval between index test and reference standard? Yes

Did all patients receive the same reference standard? Yes

Were all patients included in the analysis? Yes

Could the patient flow have introduced bias? Low risk

Iatraki 2014
Study characteristics

Patient Sampling 134 participants with subjective memory complaints were recruited from neurology clinics at the University of Crete Hospital and 3 other regional hospitals over a 20-month period.

Sampling procedure: not specified

Inclusion criteria: not specified

Exclusion criteria: not specified

Patient characteristics and setting Outpatient setting

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

Iatraki 2014 (Continued)

Study includes details of pilot study and community sample performed for normative data. The clinical sample includes 134 participants who were further divided into 3 groups; probable AD, MCI and acute neurological or psychiatric disease without clinically significant cognitive impairment.

Total number of participants: 134

Number of participants in qualitative analysis: 134

Mean age (years): 73.40±10.58 (AD), 68.19±10.00 (MCI), 65.86±13.77 (acute neurological/psychiatric condition)

Sex (% female): 42% (AD), 63% (MCI), 35% (acute neurological/psychiatric condition)

Education (years): mean 7.77±4.34 (AD), 8.67±4.19 (MCI), 8.49±4.65 (acute neurological/psychiatric condition)

Index tests	TYM-G (Greek version)
Target condition and reference standard(s)	<u>Target conditions:</u> AD and MCI <u>Reference standards:</u> NINCDS-ADRDA for probable AD, International Working Group Criteria on MCI
Flow and timing	TYM-G and clinical assessment were performed on the same day.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Iatraki 2014 (Continued)

Are there concerns that the index test, its conduct, or interpretation differ from the review question?

High

DOMAIN 3: Reference standard

Is the reference standard likely to correctly classify the target condition?

Yes

Were the reference standard results interpreted without knowledge of the results of the index tests?

Yes

Could the reference standard, its conduct, or its interpretation have introduced bias?

Low risk

Are there concerns that the target condition as defined by the reference standard does not match the question?

Low concern

DOMAIN 4: Flow and timing

Was there an appropriate interval between index test and reference standard?

Yes

Did all patients receive the same reference standard?

Yes

Were all patients included in the analysis?

Yes

Could the patient flow have introduced bias?

Low risk

Kim 2019
Study characteristics

Patient Sampling

420 participants were recruited from 11 hospitals in Korea, specifically the outpatient neurology and psychiatry clinics and regional dementia centres over a 9-month period.

Sampling procedure: not specified

Inclusion criteria: age > 65 years, clinical dementia rating score 0.5, 1 or 2

Exclusion criteria: head injury that resulted in loss of consciousness or cognitive impairment > 1 hour, previous history of intracranial haemorrhage, presence of space occupying lesion, cognitive impairment associated with neurosyphilis, HIV infection, thyroid abnormalities or vitamin B12/folate deficiency, history of metabolic encephalopathy, participants with Axis I psychiatric disorders, physical illnesses that could interfere with clinical investigations.

Patient characteristics and setting

Outpatient setting

Total number of participants: 420

Number of participants in qualitative analysis: 420

Mean age (years): 75.3±6.0

Kim 2019 (Continued)

	Sex (% female): 55.5%		
	Education (years): mean 9.1±4.9		
Index tests	KDSQ-C		
Target condition and reference standard(s)	<u>Target conditions</u> : dementia including subtypes, MCI <u>Reference standards</u> : DSM-IV for dementia, Petersen criteria for MCI, NINCDS-ADRDA		
Flow and timing	Interval between index test and reference standard was not clear.		
Comparative			
Notes			
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Kim 2019 (Continued)

Are there concerns that the target condition as defined by the reference standard does not match the question?

Low concern

DOMAIN 4: Flow and timing

Was there an appropriate interval between index test and reference standard? Unclear

Did all patients receive the same reference standard? Yes

Were all patients included in the analysis? Yes

Could the patient flow have introduced bias?

Low risk

Mansbach 2020

Study characteristics

Patient Sampling	59 participants were recruited from independent living facilities who had been referred by the attending physician for the evaluation of cognitive concerns and mood functioning over a 2-month period. <u>Sampling procedure:</u> not specified <u>Inclusion criteria:</u> not specified <u>Exclusion criteria:</u> age < 50 years, inability to complete index test without assistance
Patient characteristics and setting	Community setting 59 participants were included in Study 1 which was an exploratory analysis. Study 2 was a confirmatory analysis involving over 300 participants but this was used to assess the convergent validity of the index test and BCAT-SF (Brief-Cognitive Assessment Tool-Short Form). <u>Total number of participants:</u> 59 <u>Number of participants in qualitative analysis:</u> 59 <u>Mean age (years):</u> 80.83±9.43, range 57-97 <u>Sex (% female):</u> 66.1% <u>Education:</u> ≤ 11 years 5.08%, 12 years 32.2%, 13-15 years 10.16%, 16 years 28.81%, > 17 years 20.3%, missing 3.39%
Index tests	myMemCheck
Target condition and reference standard(s)	<u>Target conditions:</u> dementia, MCI <u>Reference standard:</u> DSM-IV
Flow and timing	The time interval between myMemCheck and DSM-IV was not reported.

Mansbach 2020 (Continued)

Comparative

Notes	Study authors contacted December 2020 for further clarification on data.
-------	--

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		

Mansbach 2020 (Continued)

Were all patients included in the analysis?

Yes

Could the patient flow have introduced bias?

Low risk

Postel-Vinay 2014
Study characteristics

Patient Sampling

201 consecutive patients with memory complaints who visited a memory clinic for the first time were recruited over a 10-month period.

Sampling procedure: consecutive

Inclusion criteria: not specified

Exclusion criteria: inability to read or write or understand French, known dementia, major depressive disorder

Patient characteristics and setting

Outpatient setting

Total number of participants: 201

Number of participants included in qualitative analysis: 201

Mean age (years): 76±10 (range 41 to 94)

Sex (% female): 68%

Education: 51% primary education, 21% secondary education, 28% higher degree

Index tests

TYM-F (French version)

Target condition and reference standard(s)

Target conditions: dementia, MCI

Reference standard: DSM-IV for dementia, European Consortium criteria and Petersen criteria for MCI

Flow and timing

TYM-F test and the clinical assessment were performed on the same day.

Comparative

Notes

Methodological quality
Item
Authors' judgement
Risk of bias
Applicability concerns
DOMAIN 1: Patient selection

Was a consecutive or random sample of patients enrolled?

Yes

Was a case-control design avoided?

Yes

Postel-Vinay 2014 (Continued)

Did the study avoid inappropriate exclusions?	Yes	
Could the selection of patients have introduced bias?		Low risk
Are there concerns that the included patients and setting do not match the review question?		Low concern
DOMAIN 2: Index test (All tests)		
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes	
If a threshold was used, was it pre-specified?	No	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 3: Reference standard		
Is the reference standard likely to correctly classify the target condition?	Yes	
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		High
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Scharre 2010
Study characteristics

Patient Sampling	254 participants were recruited from a wide variety of settings including geriatric outpatient clinics, educational talks to the lay public, independent and assisted living facilities, senior centres, free memory screens through newspaper advertisements and memory clinics over a 4-year period. Sampling procedure: not specified
------------------	---

Diagnostic test accuracy of self-administered cognitive assessment tools for dementia (Review)

Scharre 2010 (Continued)

Inclusion criteria: age > 59 years, sufficient vision and English literacy
Exclusion criteria: baseline mental retardation

Patient characteristics and setting	Community and outpatient setting Participants who completed SAGE and met inclusion criteria were divided into groups based on their SAGE scores. Participants were randomly selected from each group and asked to participate in a 1-day clinical evaluation. <u>Total number of participants</u>: 254 <u>Number of participants included in qualitative analysis</u>: 63 <u>Mean age (years)</u>: 78±7, range 60-90 years <u>Sex (% female)</u>: 66.7% <u>Education (years)</u>: mean 14.5±2.8		
Index tests	SAGE		
Target condition and reference standard(s)	<u>Target conditions</u>: dementia, MCI <u>Reference standard</u>: DSM-IV		
Flow and timing	The time interval between the initial SAGE test and the repeat SAGE test or clinical assessment was reported as 3.7 weeks (median) with a range of 0.3 to 33 weeks.		
Comparative			
Notes			
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Scharre 2010 (Continued)

Could the conduct or interpretation of the index test have introduced bias?	Low risk
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern
DOMAIN 3: Reference standard	
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern
DOMAIN 4: Flow and timing	
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Scharre 2017

Study characteristics	
Patient Sampling	426 participants who had completed SAGE at a variety of community or clinic settings over a 2-year period were eligible for the study. <u>Sampling procedure:</u> random <u>Inclusion criteria:</u> age > 50 years, sufficient vision and hearing, previously completed SAGE <u>Exclusion criteria:</u> intellectual disability, moderate to severe dementia diagnosed by a physician
Patient characteristics and setting	Community and outpatient setting Eligible participants were divided into groups depending on their SAGE scores. Participants were randomly selected from each group and asked to participate in a 1-day clinical evaluation. The study took place over a 1-year period. <u>Total number of participants:</u> 426 <u>Number of participants included in qualitative analysis:</u> 66

Scharre 2017 (Continued)

Mean age (years): 75.2+7.3, range 54-92

Sex (% female): 67%

Education (years): mean 15.1+2.7, range 8-21

Index tests	eSAGE
Target condition and reference standard(s)	<u>Target conditions</u> : dementia, MCI <u>Reference standard</u> : DSM-IV
Flow and timing	eSage and clinical assessment were completed on the same day.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Scharre 2017 (Continued)

Could the reference standard, its conduct, or its interpretation have introduced bias?

Low risk

Are there concerns that the target condition as defined by the reference standard does not match the question?

Low concern

DOMAIN 4: Flow and timing

Was there an appropriate interval between index test and reference standard? Yes

Did all patients receive the same reference standard? Yes

Were all patients included in the analysis? Yes

Could the patient flow have introduced bias?

Low risk

AD: Alzheimer's dementia

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders version 4

ICD-10: International Classification of Diseases 10th Revision

KDSQ-C: Korean Dementia Screening Questionnaire - Cognition

MCI: mild cognitive impairment

MoCA: Montreal cognitive assessment

NINCDS-ADRDA: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association

SAGE: Self-Administered Gerocognitive Exam

SD: standard deviation

TYM: test your memory

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Brown 2017	No dementia diagnosis data (confirmed with author).
Crane 2012	Only published as abstract (confirmed with author).
Dong 2019	No dementia diagnosis data.
Gallassia 2012	Only published as abstract (no reply from author).
Galvin 2007	Case control.
Jorgensen 2020	Case control.
Larner 2015	No dementia diagnosis data.
Miyamae 2016	Case control.
Saunders 2008	Only published as abstract (no reply from author).
Tsai 2018	Case control.
Van Mierlo 2017	More than 1 month between index test and reference standard.

DATA

Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 TYM < 42/50	2	240
2 TYM < 39/50	1	201
3 TYM < 36/50	1	1049
4 SAGE	2	84
5 CCT	1	0
6 BrainCheck Memory	1	18
7 p-KDSQ-C	1	370
8 p-KDSQ-C-I	1	370
9 p-KDSQ-C-II	1	370
10 p-KDSQ-C-III	1	370
11 p-KDSQ-C-IV	1	370
12 myMemCheck	1	0
13 TYM < 30/50	1	224
14 TYM < 31/50	1	134
15 TYM < 34/50	1	1049
16 TYM < 35/50	1	1049
17 TYM < 37/50	1	1049
18 TYM < 38/50	1	1049

APPENDICES

Appendix 1. Tabulation of study results

Study ID	Comparison	Country	Number included in qualitative analysis	Number with target condition	Index test	Maximum score	Test threshold	Sensitivity (%)	Specificity (%)				
Ball 2001	AD vs no dementia	USA	53	Not specified	CCT		Abnormal CCT score	67	69				
	AD vs no dementia		53	Not specified	CCT		Abnormal CCT score plus 2 risk factors for AD	75	87				
Breitve 2015	AD vs no dementia	Norway	16	7	TYM-N	50	42	100	56				
	Dementia vs no dementia		16	10 (7 AD, 3 either vascular or PD)	TYM-N		42	100	83				
Ferrer-Arias 2016	Dementia vs no dementia	Spain	1049	127	TYM-S	50	≤ 34	82	93				
	≤ 35						89	91					
	≤ 36						94	89					
	≤ 37						97	86					
	≤ 38						98	82					
	Cognitive impairment vs no cognitive impairment						1049	310 (127 dementia, 183 MCI)	TYM-S	50	≤ 38	75	96
	≤ 39						80	93					
						≤ 40	86	88					

(Continued)

							≤ 41	89	84
							≤ 42	93	77
Groppell 2019	Cognitive impairment vs no cognitive impairment	USA	18	9	BrainCheck Memory	30	Not specified	89	78
Hancock 2010	Dementia vs no dementia	UK	224	78	TYM	50	≤ 42	95	45
							≤ 30	73	88
Iatraki 2014	AD vs no dementia	Greece	134	43	TYM-G	50	31	81	71
							Other test thresholds 26, 42 and 43 but this is normative data taken from the community sample		
Kim 2019	Dementia vs no dementia	Korea	420	170	KDSQ	Not specified			
					p-KDSQ-C		< 6	62	77
					p-KDSQ-C-I		< 3	64	74
					p-KDSQ-C-II		< 3	73	71
					p-KDSQ-C-III		< 6	55	80
					p-KDSQ-C-IV		< 3	66	68
Mansbach 2020	Dementia vs no dementia	USA	59	23	MyMem-Check	8	Dementia mean score = 3.91, SD = 1.79, no dementia mean score = 5.25, SD = 1.90, P = 0.01	Not specified	Not specified

(Continued)

Pos- tel-Vinay 2014	Dementia vs no dementia	France	201	68	TYM-F	50	≤ 39	90	70
	Dementia vs no dementia (MCI excluded)						≤ 40	93	76
Scharre 2010	Dementia vs no dementia (MCI excluded)	USA	63	21	SAGE	22	≤ 16	95	95
	Dementia vs no dementia						≤ 14	81	88
Scharre 2017	Dementia vs no dementia (MCI excluded)	USA	66	21	e-SAGE	22	≤ 14	95	100
	Dementia vs no dementia						≤ 13	90	87

AD: Alzheimer's disease; MCI: mild cognitive impairment; PD: Parkinson's disease dementia; SD: standard deviation; vs: versus

Index tests

CCT: Clock Completion Test

e-SAGE: electronic Self-Administered Gerocognitive Exam

KDSQ: Korean Dementia Screening Questionnaire

p-KDSQ-C: patient-rated Korean Dementia Screening Questionnaire-Cognition

p-KDSQ-C-I: patient-rated Korean Dementia Screening Questionnaire-Cognition items 1-5 memory domain

p-KDSQ-C-II: patient-rated Korean Dementia Screening Questionnaire-Cognition items 1-5 & 11-15, memory domain + activities of daily living

p-KDSQ-C-III: patient-rated Korean Dementia Screening Questionnaire-Cognition items 1-5 & 6-10, memory domain + other cognitive domains

p-KDSQ-C-IV: patient-rated Korean Dementia Screening Questionnaire-Cognition items 6-10 & 11-15, other cognitive domains + activities of daily living

SAGE: Self-Administered Gerocognitive Exam

TYM: Test Your Memory

TYM-F: Test Your Memory- French

TYM-G: Test Your Memory-Greek

TYM-N: Test Your Memory-Norwegian

TYM-S: Test Your Memory-Spanish

Appendix 2. Sources searched and search strategies

Source	Search strategy	Hits retrieved
1. MEDLINE In-process and other non-indexed citations and MEDLINE 1950-present (Ovid SP) (Date of most recent search: 2 November 2022)	1 exp Dementia/	Aug 2020: 756
	2 Delirium/	June 2021: 98
	3 Wernicke Encephalopathy/	April 2022: 131
	4 Delirium, Dementia, Amnestic, Cognitive Disorders/	Nov 2022: 61
	5 dement*.tw,kw.	
	6 alzheimer*.tw,kw.	
	7 (lewy* adj2 bod*).tw,kw.	
	8 (chronic adj2 cerebrovascular).tw,kw.	
	9 ("organic brain disease" or "organic brain syndrome").tw,kf.	
	10 "benign senescent forgetfulness".tw,kf.	
	11 (cerebr* adj2 deteriorat*).tw,kf.	
	12 (cerebral* adj2 insufficient*).tw,kf.	
	13 "major neurocognitive disorder*".tw,kf.	
	14 or/1-13	
	15 "self asses*".tw,kf.	
	16 "self test*".tw,kf.	
	17 "self admin*".tw,kf.	
	18 "self complet*".tw,kf.	
	19 "test your memory".tw,kf.	

(Continued)

- 20 TYM.tw,kf.
- 21 "TYM-MCI".tw,kf.
- 22 "TYM-TR".tw,kf.
- 23 "F-TYM".tw,kf.
- 24 "TYM-S".tw,kf.
- 25 "TYM-CN".tw,kf.
- 26 "TYM-HUN".tw,kf.
- 27 "TYM-J".tw,kf.
- 28 eSAGE.tw,kf.
- 29 (SAGE and (test* or exam*)).tw,kf.
- 30 "Brainlab Cognition".tw,kf.
- 31 "DANA Modular".tw,kf.
- 32 "DANAe Brain Vital".tw,kf.
- 33 "Dementia Screener".tw,kf.
- 34 "Dementia Test Risk Calculator of Dementia".tw,kf.
- 35 "Imprint Memory Assessment".tw,kf.
- 36 "MyBrainTest".tw,kf.
- 37 "Psychology Today Memory Test".tw,kf.
- 38 "Savonix Mobile".tw,kf.
- 39 "The Cleveland Clinic Brain Check Up".tw,kf.
- 40 "The Online Memory Screening Test".tw,kf.
- 41 BrainCheck.tw,kf.
- 42 BrainTest.tw,kf.
- 43 Cogniciti.tw,kf.
- 44 Husketest.tw,kf.
- 45 MemTrax.tw,kf.
- 46 Mindcrowd.tw,kf.
- 47 RateMyMemory.tw,kf.
- 48 myMEMCheck.tw,kf.
- 49 exp "Mental Status and Dementia Tests"/ and self*.tw,kf.
- 50 or/15-49
- 51 14 and 50
- 52 exp animals/ not humans.sh.
- 53 51 not 52

(Continued)

2. Embase via Ovid	1 exp dementia/	Aug 2020: 1403
1974 to present	2 delirium/	June 2021: 195
(Date of most recent search: 2 November 2022)	3 Wernicke encephalopathy/	April 2022: 228
	4 dement*.tw,kw.	Nov 2022: 171
	5 alzheimer*.tw,kw.	
	6 (lewy* adj2 bod*).tw,kw.	
	7 (chronic adj2 cerebrovascular).tw,kw.	
	8 ("organic brain disease" or "organic brain syndrome").tw,kw.	
	9 "benign senescent forgetfulness".tw,kw.	
	10 (cerebr* adj2 deteriorat*).tw,kw.	
	11 (cerebral* adj2 insufficient*).tw,kw.	
	12 "major neurocognitive disorder*".tw,kw.	
	13 or/1-12	
	14 "self asses*".tw,kw.	
	15 "self test*".tw,kw.	
	16 "self admin*".tw,kw.	
	17 "self complet*".tw,kw.	
	18 "test your memory".tw,kw.	
	19 TYM.tw,kw.	
	20 "TYM-MCI".tw,kw.	
	21 "TYM-TR".tw,kw.	
	22 "F-TYM".tw,kw.	
	23 "TYM-S".tw,kw.	
	24 "TYM-CN".tw,kw.	
	25 "TYM-HUN".tw,kw.	
	26 "TYM-J".tw,kw.	
	27 eSAGE.tw,kw.	
	28 (SAGE and (test* or exam*)).tw,kw.	
	29 "Brainlab Cognition".tw,kw.	
	30 "DANA Modular".tw,kw.	
	31 "DANAe Brain Vital".tw,kw.	
	32 "Dementia Screener".tw,kw.	
	33 "Dementia Test Risk Calculator of Dementia".tw,kw.	
	34 "Imprint Memory Assessment".tw,kw.	

(Continued)

- 35 "MyBrainTest".tw,kw.
- 36 "Psychology Today Memory Test".tw,kw.
- 37 "Savonix Mobile".tw,kw.
- 38 "The Cleveland Clinic Brain Check Up".tw,kw.
- 39 "The Online Memory Screening Test".tw,kw.
- 40 BrainCheck.tw,kw.
- 41 BrainTest.tw,kw.
- 42 Cogniciti.tw,kw.
- 43 Husketest.tw,kw.
- 44 MemTrax.tw,kw.
- 45 Mindcrowd.tw,kw.
- 46 RateMyMemory.tw,kw.
- 47 myMEMCheck.tw,kw.
- 48 exp "diagnostic test"/ and self*.tw,kw.
- 49 or/14-48
- 50 13 and 49

3. PSYCINFO via Ovid (Date of most recent search: 2 November 2022)	1 exp Dementia/	Aug 2020: 526
	2 exp Delirium/	June 2021: 65
	3 exp Huntingtons Disease/	April 2022: 76
	4 exp Kluver Bucy Syndrome/	Nov 2022: 49
	5 exp Wernickes Syndrome/	
	6 exp Cognitive Impairment/	
	7 dement*.mp.	
	8 alzheimer*.mp.	
	9 (lewy* adj2 bod*).mp.	
	10 deliri*.mp.	
	11 (chronic adj2 cerebrovascular).mp.	
	12 ("organic brain disease" or "organic brain syndrome").mp.	
	13 "supranuclear palsy".mp.	
	14 ("normal pressure hydrocephalus" and "shunt*").mp.	
	15 "benign senescent forgetfulness".mp.	
	16 (cerebr* adj2 deteriorat*).mp.	
	17 (cerebral* adj2 insufficient*).mp.	
	18 (pick* adj2 disease).mp.	

(Continued)

- 19 (creutzfeldt or jcd or cjd).mp.
- 20 huntington*.mp.
- 21 binswanger*.mp.
- 22 korsako*.mp.
- 23 ("parkinson* disease dementia" or PDD or "parkinson* dementia").mp.
- 24 "major neurocognitive disorder".ti,ab.
- 25 or/1-24
- 26 "self asses*".ti,ab.
- 27 "self test*".ti,ab.
- 28 "self admin*".ti,ab.
- 29 "self complet*".ti,ab.
- 30 "test your memory".ti,ab.
- 31 TYM.ti,ab.
- 32 "TYM-MCI".ti,ab.
- 33 "TYM-TR".ti,ab.
- 34 "F-TYM".ti,ab.
- 35 "TYM-S".ti,ab.
- 36 "TYM-CN".ti,ab.
- 37 "TYM-HUN".ti,ab.
- 38 "TYM-J".ti,ab.
- 39 eSAGE.ti,ab.
- 40 (SAGE and (test* or exam*)).ti,ab.
- 41 "Brainlab Cognition".ti,ab.
- 42 "DANA Modular".ti,ab.
- 43 "DANAe Brain Vital".ti,ab.
- 44 "Dementia Screener".ti,ab.
- 45 "Dementia Test Risk Calculator of Dementia".ti,ab.
- 46 "Imprint Memory Assessment".ti,ab.
- 47 "MyBrainTest".ti,ab.
- 48 "Psychology Today Memory Test".ti,ab.
- 49 "Savonix Mobile".ti,ab.
- 50 "The Cleveland Clinic Brain Check Up".ti,ab.
- 51 "The Online Memory Screening Test".ti,ab.
- 52 BrainCheck.ti,ab.

(Continued)

53 BrainTest.ti,ab.
54 Cogniciti.ti,ab.
55 Husketest.ti,ab.
56 MemTrax.ti,ab.
57 Mindcrowd.ti,ab.
58 RateMyMemory.ti,ab.
59 myMEMCheck.ti,ab.
60 exp Cognitive Assessment/ and self*.ti,ab.
61 or/26-60
62 25 and 61

4. Web of Science – BIOSIS Citation Index (Date of most recent search: 2 November 2022)	(dement* OR alzheimer* OR "vascular cognitive impairment" OR "lew* bod*" OR CADASIL OR "cognit* impair*" OR FTD OF FTLD OR "cerebrovascular insufficienc*" OR AD OR VCI) AND TOPIC: ("self asses*" OR "self test*" OR "self admin*" OR "self complet*" OR TYM OR SAGE)	Aug 2020: 668
		June 2021: 226
		April 2022: 408
		Nov 2022: 48
5. LILACS (BIREME) (Date of most recent search: 2 November 2022)	dement\$ OR Alzheimer\$ OR "vascular cognitive impairment" OR "lew\$ bod\$" OR CADASIL OR "cognit\$ impair\$" OR FTD OF FTLD OR "cerebrovascular insufficienc\$" OR AD OR VCI [Words] and "self asses\$" OR "self test\$" OR "self admin\$" OR "self complet\$" OR TYM OR SAGE [Words]	AUG 2020: 30
		June 2021: 1
		April 2022: 1
		Nov 2022: 0
6. CINAHL (EBSCO) (Date of most recent search: 2 November 2022)	S57 S20 AND S56	Aug 2020: N/A
	S56 S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55	June 2021: 666
	S55 TX "sage adj3 exam*"	April 2022: 68
	S54 TX "sage adj3 test*"	Nov 2022: 34
	S53 TX myMEMCheck	
	S52 TX RateMyMemory	
	S51 TX Mindcrowd	
	S50 TX MemTrax	
	S49 TX Husketest	
	S48 TX Cogniciti	
	S47 TX BrainTest	
	S46 TX BrainCheck	
	S45 TX "The Online Memory Screening Test"	
	S44 TX "The Cleveland Clinic Brain Check Up"	

(Continued)

S43 TX "Savonix Mobile"

S42 TX "Psychology Today Memory Test"

S41 TX "MyBrainTest"

S40 TX "Imprint Memory Assessment"

S39 TX "Dementia Test Risk Calculator of Dementia"

S38 TX "Dementia Screener"

S37 TX "DANAe Brain Vital"

S36 TX "DANA Modular"

S35 TX "Brainlab Cognition"

S34 TX eSAGE

S33 TX TYM-J

S32 TX TYM-HUN

S31 TX TYM-CN

S30 TX TYM-S

S29 TX F-TYM

S28 TX TYM-TR

S27 TX TYM-MCI

S26 TX TYM

S25 TX "test your memory"

S24 TX "self complet**"

S23 TX "self admin**"

S22 TX "self test**"

S21 TX "self asses**"

S20 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR
S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19

S19 TX "major neurocognitive disorder"

S18 TX korsako*

S17 TX binswanger*

S16 TX huntington*

S15 TX creutzfeldt or jcd or cjd

S14 TX pick* N2 disease

S13 TX cerebral* N2 insufficient*

S12 TX cerebr* N2 deteriorat*

S11 TX "benign senescent forgetfulness"

S10 TX "normal pressure hydrocephalus" and "shunt**"

(Continued)

S9 TX "organic brain disease" or "organic brain syndrome"

S8 TX chronic N2 cerebrovascular

S7 TX deliri*

S6 TX lewy* N2 bod*

S5 TX alzheimer*

S4 TX dement*

S3 MH "Wernicke's Encephalopathy"

S2 (MH "Delirium") or (MH "Delirium, Dementia, Amnestic, Cognitive Disorders")

S1 MH "Dementia+"

Aug 2020: 3383

June 2021: 1251

April 2022: 912

Nov 2022: 363

Total: 5909

Aug 2020: 2446

June 2021: 674

April 2022: 466

Nov 2022: 277

Total: 3863

Appendix 3. QUADAS-2 assessment

We provide some core anchoring statements for quality assessment of diagnostic test accuracy reviews of neuropsychological tests in dementia. These statements are designed for use with the QUADAS-2 tool and were derived during a two-day, multidisciplinary focus group in 2010. If a QUADAS-2 signalling question for a specific domain is answered 'yes', the risk of bias can be judged to be 'low'. If a question is answered 'no', this indicates a risk of potential bias. The focus group was tasked with judging the extent of the bias for each domain. During this process, it became clear that certain issues were key to assessing quality, whilst others were important to record but were less important for assessing overall quality. To assist, we describe a 'weighting' system. When an item is weighted 'high risk', that section of the QUADAS-2 results table is judged to have a high potential for bias if a signalling question is answered 'no'. For example, in dementia diagnostic test accuracy studies, ensuring that clinicians performing dementia assessment are blinded to results of the index test is fundamental. If this blinding was not present, the item on the reference standard should be scored 'high risk of bias', regardless of the other contributory elements. When an item is weighted 'low risk', it is judged to have a low potential for bias if a signalling question for that section of the QUADAS-2 results table is answered 'no'. Overall bias is judged on whether other signalling questions (with a high risk of bias) for the same domain are also answered 'no'. In assessing individual items, a score of 'unclear' should be given only if there is genuine uncertainty. In these situations, the review authors contact the relevant study teams for additional information.

Anchoring statements to assist with risk of bias assessment

Domain 1: patient selection

Risk of bias: could the selection of patients have introduced bias? (high/low/unclear)

Was a consecutive or random sample of patients enrolled? When sampling is used, the methods least likely to cause bias are consecutive sampling and random sampling, which should be stated or described, or both. Non-random sampling or sampling based on volunteers is more likely to be at high risk of bias. **Weighting: high risk of bias**

Was a case-control design avoided? Case-control study designs have a high risk of bias, but are sometimes the only studies available, especially if the index test is expensive or invasive, or both. Nested case-control designs (systematically selected from a defined population cohort) are less prone to bias, but they will still narrow the spectrum of patients that receive the index test. Study designs (both cohort and case-control) that may also increase bias are those designs in which the study team deliberately increases or decreases the proportion of participants with the target condition, for example a population study may be enriched with extra dementia participants from a secondary care setting. **Weighting: high risk of bias**

Did the study avoid inappropriate exclusions? The study will be automatically graded as unclear if exclusions are not detailed (pending contact with study authors). When exclusions are detailed, we graded the study as 'low risk' if we felt that the exclusions were appropriate. Certain exclusions common to many studies of dementia are medical instability, terminal disease, alcohol/substance misuse, concomitant psychiatric diagnosis, and other neurodegenerative condition. However, if 'difficult-to-diagnose' groups are excluded, this may introduce bias, so exclusion criteria must be justified. For a community sample, we would expect relatively few exclusions. We label post hoc exclusions as 'high risk' of bias. **Weighting: high risk of bias**

Applicability: are there concerns that the included patients do not match the review question? (high/low/unclear)

The included patients should match the intended population as described in the review question. If not already specified in the review inclusion criteria, the setting will be particularly important – the review authors should consider population in terms of symptoms, pretesting, and potential disease prevalence. We will classify studies that use very selected participants or subgroups as low applicability, unless they are intended to represent a defined target population, for example people with memory problems referred to a specialist and investigated by lumbar puncture.

Domain 2: Index test

Risk of bias: could the conduct or interpretation of the index test have introduced bias? (high/low/unclear)

Were the index test results interpreted without knowledge of the reference standard? Terms such as 'blinded' or 'independently and without knowledge of' are sufficient; full details of the blinding procedure are not required. This item may be scored as 'low risk' if it is explicitly described, or if there is a clear temporal pattern to the order of testing that precludes the need for formal blinding (e.g. all (neuropsychological test) assessments were performed before the dementia assessment). As most neuropsychological tests are administered by a third party, knowledge of dementia diagnosis may influence their ratings; tests that are self-administered, for example by using a computerised version, may have less risk of bias. **Weighting: high risk**

Were the index test thresholds prespecified? For neuropsychological scales, there is usually a threshold above which participants are classified as 'test positive'; this may be referred to as threshold, clinical cut-off, or dichotomisation point. Different thresholds are used in different populations. A study is classified as at higher risk of bias if the authors define the optimal cut-off post hoc based on their own study data. Certain papers may use an alternative methodology for analysis that does not use thresholds; these papers should be classified as not applicable. **Weighting: low risk**

Were sufficient data on (neuropsychological test) application given for the test to be repeated in an independent study? Particular points of interest include method of administration (e.g. self-completed questionnaire versus direct-questioning interview), nature of informant, and language of assessment. If a novel form of the index test is used, for example a translated questionnaire, details of the scale should be included and a reference given to an appropriate descriptive text, and evidence of validation should be provided. **Weighting: low risk**

Applicability: are there concerns that the index test, its conduct, or its interpretation may differ from the review question? (high/low/unclear)

Variations in the length, structure, language, and/or administration of the index test may all affect applicability if they differ from those specified in the review question.

Domain 3: Reference standard

Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias? (high/low/unclear)

Is the reference standard likely to correctly classify the target condition? Commonly used international criteria that can assist with clinical diagnosis of dementia include those detailed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and the International Classification of Diseases (ICD-10). Criteria specific to dementia subtypes include but are not limited to the National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for Alzheimer's dementia; McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; and National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria for vascular dementia. When the criteria used for assessment are unfamiliar to the review authors and the Cochrane Dementia and Cognitive Improvement Group, this item should be classified as 'high risk of bias'. **Weighting: high risk**

Were the reference standard results interpreted without knowledge of the results of the index test? Terms such as 'blinded' or 'independent' are sufficient; full details of the blinding procedure are not required. This may be scored as 'low risk' if explicitly described, or if a clear temporal pattern to the order of testing is evident (e.g. all dementia assessments performed before (neuropsychological test) testing). Informant rating scales and direct cognitive tests present certain problems. It is accepted that informant interview and cognitive

testing are usual components of clinical assessment for dementia, however, specific use of the scale under review in the clinical dementia assessment should be scored as high risk of bias. **Weighting: high risk**

Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study? Particular points of interest for dementia assessment include the training or expertise of the assessor; whether additional information (e.g. neuroimaging; other neuropsychological test results) was available to inform the diagnosis; and whether this was available for all participants. **Weighting: variable risk, but high risk if method of dementia assessment not described**

Applicability: are there concerns that the target condition as defined by the reference standard does not match the review question? (high/low/unclear)

There exists the possibility that some methods of dementia assessment, although valid, may diagnose a smaller or larger proportion of participants with disease than in usual clinical practice. In these instances, the item should be rated 'poor applicability'.

Domain 4: Patient flow and timing (N.B. refer to, or construct, a flow diagram)

Risk of bias: could the patient flow have introduced bias? (high/low/unclear)

Was there an appropriate interval between the index test and the reference standard? For a cross-sectional study design, the potential exists for the participant to change between assessments, however, dementia is a slowly progressive disease that is not reversible. The ideal scenario would be a same-day assessment, but longer periods of time (e.g. several weeks or months) are unlikely to lead to a high risk of bias. For delayed-verification studies, the index and reference tests are necessarily separated in time, given the nature of the condition. **Weighting: low risk**

Did all participants receive the same reference standard? In some scenarios, participants who score 'test positive' on the index test have a more detailed assessment for the target condition. When dementia assessment (or the reference standard) differs between participants, this should be classified as high risk of bias. **Weighting: high risk**

Were all participants included in the final analysis? Attrition will vary with study design. Delayed-verification studies will have higher attrition than cross-sectional studies because of mortality, and this is likely to be greater in participants with the target condition. Dropouts (and missing data) should be accounted for. Attrition that is higher than expected (compared with other similar studies) should be treated as high risk of bias. We have defined an attrition cut-off of more than 20% as being high risk, but this will be highly dependent on the length of follow-up in individual studies. **Weighting: high risk**

HISTORY

Protocol first published: Issue 9, 2020

CONTRIBUTIONS OF AUTHORS

F Naeem: investigation, data curation, formal analysis, writing (original draft)

J McCleery: conceptualisation, investigation, methodology, writing (review and editing)

TM Hietamies: investigation, data curation, writing (original draft)

F Abakar Ismail: investigation, data curation, writing (original draft)

S Clinton: investigation, data curation, writing (original draft)

A O'Mahony: investigation, data curation, writing (original draft)

OJ Ponce: investigation, data curation, writing (original draft)

TJ Quinn: conceptualisation, investigation, methodology, writing (review and editing)

DECLARATIONS OF INTEREST

F Naeem: none known

J McCleery: none known

TM Hietamies: none known

F Abakar Ismail: none known

S Clinton: none known

A O'Mahony: none known

OJ Ponce: none known

TJ Quinn: has a role as a Cochrane Editor but has not been involved in the editorial process for this review.

SOURCES OF SUPPORT

Internal sources

- New Source of support, Other

No sources of support identified.

External sources

- National Institute for Health and Care Research (NIHR), UK

This review were supported by the NIHR, via Cochrane Infrastructure funding to the Cochrane Dementia and Cognitive Improvement Group until 31 March 2023. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Evidence Synthesis Programme, NIHR, UK National Health Service, or the Department of Health and Social Care.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We have changed the title of the review to reflect the heterogeneity in index tests and, rather than describing them as 'questionnaires', have adopted the term cognitive assessment 'tools' as a more inclusive description as per peer review comments.

In the protocol, our intention was to look at only dementia as an outcome. For some studies we were not able to exclude data on mild cognitive impairment (MCI). Where possible, data have been reported with exclusion of MCI.

We intended to include the US National Institutes of Health Ongoing Trials Register (clinicaltrials.gov) but this was not a part of the final search strategy and results.

We did not perform all the quantitative meta-analyses outlined in the protocol due to an insufficient number of studies reporting test accuracy data for the same test at the same threshold.

In the protocol, we planned to exclude studies where the interval between the index test and the reference standard was longer than one month. Due to a lack of available studies, we have included studies with a longer or undefined test interval for some participants, but have discussed this as a potential source of heterogeneity amongst the included studies in the [Findings](#) section.

A O'Mahony, OJ Ponce, S Clinton, and F Naeem joined the review team.

INDEX TERMS

Medical Subject Headings (MeSH)

Cognition; *Dementia [diagnosis]; Mental Status and Dementia Tests [standards]; Neuropsychological Tests [standards]; *Sensitivity and Specificity

MeSH check words

Aged; Humans