

Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias (Review)

Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S



**THE COCHRANE
COLLABORATION®**

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2015, Issue 10

<http://www.thecochranelibrary.com>

WILEY

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
BACKGROUND	3
Figure 1.	4
OBJECTIVES	5
METHODS	5
Figure 2.	8
RESULTS	9
Figure 3.	10
Figure 4.	11
Figure 5.	12
Figure 6.	14
Figure 7.	14
DISCUSSION	18
AUTHORS' CONCLUSIONS	19
ACKNOWLEDGEMENTS	19
REFERENCES	20
CHARACTERISTICS OF STUDIES	23
DATA	39
Test 1. MoCA Secondary care.	39
Test 2. MoCA community based studies.	40
APPENDICES	40
CONTRIBUTIONS OF AUTHORS	49
DECLARATIONS OF INTEREST	49
SOURCES OF SUPPORT	49
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	49

[Diagnostic Test Accuracy Review]

Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias

Daniel HJ Davis¹, Sam T Creavin², Jennifer LY Yip³, Anna H Noel-Storr⁴, Carol Brayne³, Sarah Cullum²

¹MRC Unit for Lifelong Health and Ageing, University College London, London, UK. ²School of Social and Community Medicine, University of Bristol, Bristol, UK. ³Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁴Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Contact address: Daniel HJ Davis, MRC Unit for Lifelong Health and Ageing, University College London, 33 Bedford Place, London, WC1B 5JU, UK. daniel.davis@ucl.ac.uk.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.

Publication status and date: New, published in Issue 10, 2015.

Review content assessed as up-to-date: 2 August 2012.

Citation: Davis DHJ, Creavin ST, Yip JLY, 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.

Copyright © 2015 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Dementia is a progressive syndrome of global cognitive impairment with significant health and social care costs. Global prevalence is projected to increase, particularly in resource-limited settings. Recent policy changes in Western countries to increase detection mandates a careful examination of the diagnostic accuracy of neuropsychological tests for dementia.

Objectives

To determine the diagnostic accuracy of the Montreal Cognitive Assessment (MoCA) at various thresholds for dementia and its subtypes.

Search methods

We searched MEDLINE, EMBASE, BIOSIS Previews, Science Citation Index, PsycINFO and LILACS databases to August 2012. In addition, we searched specialised sources containing diagnostic studies and reviews, including MEDION (Meta-analyses van Diagnostisch Onderzoek), DARE (Database of Abstracts of Reviews of Effects), HTA (Health Technology Assessment Database), ARIF (Aggressive Research Intelligence Facility) and C-EBLM (International Federation of Clinical Chemistry and Laboratory Medicine Committee for Evidence-based Laboratory Medicine) databases. We also searched ALOIS (Cochrane Dementia and Cognitive Improvement Group specialized register of diagnostic and intervention studies). We identified further relevant studies from the PubMed 'related articles' feature and by tracking key studies in Science Citation Index and Scopus. We also searched for relevant grey literature from the Web of Science Core Collection, including Science Citation Index and Conference Proceedings Citation Index (Thomson Reuters Web of Science), PhD theses and contacted researchers with potential relevant data.

Selection criteria

Cross-sectional designs where all participants were recruited from the same sample were sought; case-control studies were excluded due to high chance of bias. We searched for studies from memory clinics, hospital clinics, primary care and community populations. We excluded studies of early onset dementia, dementia from a secondary cause, or studies where participants were selected on the basis of a specific disease type such as Parkinson's disease or specific settings such as nursing homes.

Data collection and analysis

We extracted dementia study prevalence and dichotomised test positive/test negative results with thresholds used to diagnose dementia. This allowed calculation of sensitivity and specificity if not already reported in the study. Study authors were contacted where there was insufficient information to complete the 2x2 tables. We performed quality assessment according to the QUADAS-2 criteria.

Methodological variation in selected studies precluded quantitative meta-analysis, therefore results from individual studies were presented with a narrative synthesis.

Main results

Seven studies were selected: three in memory clinics, two in hospital clinics, none in primary care and two in population-derived samples. There were 9422 participants in total, but most of studies recruited only small samples, with only one having more than 350 participants. The prevalence of dementia was 22% to 54% in the clinic-based studies, and 5% to 10% in population samples. In the four studies that used the recommended threshold score of 26 or over indicating normal cognition, the MoCA had high sensitivity of 0.94 or more but low specificity of 0.60 or less.

Authors' conclusions

The overall quality and quantity of information is insufficient to make recommendations on the clinical utility of MoCA for detecting dementia in different settings. Further studies that do not recruit participants based on diagnoses already present (case-control design) but apply diagnostic tests and reference standards prospectively are required. Methodological clarity could be improved in subsequent DTA studies of MoCA by reporting findings using recommended guidelines (e.g. STARDdem). Thresholds lower than 26 are likely to be more useful for optimal diagnostic accuracy of MoCA in dementia, but this requires confirmation in further studies.

PLAIN LANGUAGE SUMMARY

Montreal Cognitive Assessment (MoCA) for the diagnosis of Alzheimer's disease and other dementias

Review question

We reviewed the evidence about the accuracy of the Montreal Cognitive Assessment (MoCA) test for diagnosing dementia and its subtypes.

Background

Dementia is a common condition in older people, with at least 7% of people over 65 years old in the UK affected, and numbers are increasing worldwide. In this review, we wanted to discover whether using a well-established cognitive test, MoCA, could accurately detect dementia when compared to a gold standard diagnostic test. MoCA uses a series of questions to test different aspects of mental functioning.

Study characteristics

The evidence we reviewed is current to August 2012. We found seven studies that matched our criteria. There were three from memory clinics (specialist clinics where people are referred for suspected dementia), two from general hospital clinics, none from primary care and two studies carried out in the general population. All studies included older people, with the youngest average age of 61 years in one study. There were a total 9422 people included in all 7 studies though only one study had more than 350 people.

The proportion of people with dementia was 5% to 10% in two population-derived studies and 22% to 54% in the five clinic-based studies. There was a large variation in the way the different studies were carried out: therefore we chose to present the results in a narrative summary because a statistical summary (combining all the estimates into a summary sensitivity and specificity) would not have been meaningful.

Key results

We found that MoCA was good at detecting dementia when using a recognised cut-off score of less than 26. In the studies that used this cut-off, we found the test correctly detected over 94% of people with dementia in all settings. On the other hand, the test also produced a high proportion of false positives, that is people who did not have dementia but tested positive at the 'less than 26' cut-off. In the studies we reviewed, over 40% of people without dementia would have been incorrectly diagnosed with dementia using the MoCA.

Conclusion

The overall quality of the studies was not good enough to make firm recommendations on using the MoCA to detect dementia in different healthcare settings. In particular, no studies looked at how useful MoCA is for diagnosing dementia in primary care settings. It is likely that a MoCA threshold lower than 26 would be more useful for optimal diagnostic accuracy in dementia, though this requires wider confirmation.

BACKGROUND

Target condition being diagnosed

The target condition in this review is dementia (all-cause) and any common dementia subtype. Dementia is a progressive syndrome of global cognitive impairment, affecting at least 7% of the UK population over 65 years of age (Alzheimer's Society 2014). World-wide, 36 million people were estimated to be living with dementia in 2010, and this number will increase to more than 115 million by 2050, with the greatest increases in prevalence predicted to occur in resource-limited settings (Wimo 2010). Dementia encompasses a group of disorders characterised by progressive loss of both cognitive function and ability to perform activities of daily living and can be accompanied by neuropsychiatric symptoms and challenging behaviours of varying type and severity. The underlying pathology is usually degenerative, and subtypes of dementia include Alzheimer's disease dementia (ADD), vascular dementia, dementia with Lewy bodies, and frontotemporal dementia.

Index test(s)

The index test is the Montreal Cognitive Assessment (MoCA), which is a measure of global cognitive function (Nasreddine 2005). It was originally developed to detect mild cognitive impairment (MCI) but is now frequently used as a screening tool for the dementias (Alzheimer's Society 2013). The MoCA is scored out of 30 points. The raw score is adjusted by educational attainment (1 extra point for 10 to 12 years of formal education; 2 points added for 4 to 9 years of formal education). The original study gave the normal range for MoCA as 26 to 30 points (Nasreddine 2005); the suggested threshold of 25/26 was the same for MCI and for dementia (scoring 25 or below being indicative of cognitive

impairment). This cut-point is now widely used as the threshold for detecting cognitive impairment and possible dementia. The MoCA is recommended by the Alzheimer's Society as one of the tests that can be used for detection of dementia in memory clinics in the UK (Alzheimer's Society 2013).

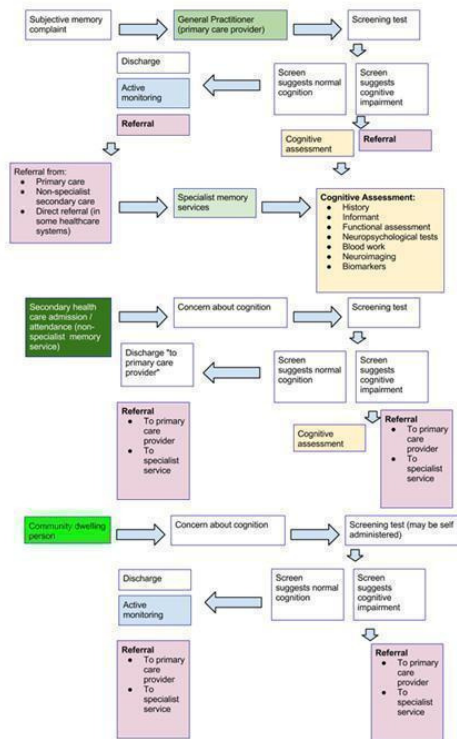
The MoCA is a brief test of cognitive function, taking 10 minutes to administer (Ismail 2010). It assesses short-term memory, visuospatial function, executive function, attention, concentration and working memory, language, and orientation. Compared to the Mini-Mental State Examination (MMSE) (Folstein 1975), the MoCA offers more detailed testing of executive function (Nasreddine 2005). Recently, the MoCA has been considered as an alternative to the MMSE as a screening test for non-specialist use since the latter is now copyrighted and there is a charge for its use. However, the MoCA is the more difficult test, so scores have never been regarded as being equivalent.

Three versions of the MoCA exist in English to minimise practice effects. Each version tests the same domains, but the content of specific tasks differ (e.g. different words for episodic memory, different pictures for semantic memory). These alternative versions are reported to have comparable reliability to the original (Costa 2012). Construct validity has also been assessed and is concordant with standard neuropsychological testing (Freitas 2012). Multiple translations are also available, as is a version for visually impaired persons (Wittich 2010). These are all published online (www.mocatest.org).

Clinical pathway

Dementia usually develops over several years. Individuals, or their relatives, may notice subtle impairments of recent memory. Gradually, more cognitive domains become involved, and difficulty in planning complex tasks becomes increasingly apparent. Figure 1 gives an overview of routes through which individuals may present.

Figure 1. Diagnostic pathways in dementia



Notes

1. Some persons with a normal cognitive screen may still be referred on due to individual patient or doctor factors (for example clinical uncertainty, unusual features or high health anxiety)
2. A cognitive assessment leading to a formal diagnosis may be performed by primary healthcare services, secondary (non-specialist memory clinic) services or specialised memory clinic services but is most likely to occur in the context of a service which is designed to provide a formal diagnosis of a dementia or dementia subtype (that is a memory clinic)
3. Secondary care may refer to primary care in some healthcare settings if the funding follows the patient referral or if inter-disciplinary referrals are not permitted or are un-funded by the health care purchaser
4. Community dwelling persons may have a cognitive test initiated by themselves (if they have concerns about their memory) or by others (e.g. carers, relatives). They will be unable to perform a cognitive assessment as they have no clinical training.

The pathway to dementia diagnosis influences the diagnostic test accuracy of the MoCA, and thus the aim of this review is to separate the analyses by study population. Presenting the findings of the review by study population emphasises that the utility of the test differs across settings, and that guidance is needed to decide where and how the test would best be used (Brayne 2012).

The MoCA was developed in 2005 and evidence for its diagnostic utility in different settings is continuing to emerge. Often the first diagnostic accuracy studies for new tests are conducted in specialist clinics, but estimates of the utility of the MoCA are also needed for other groups: hospital inpatients; general outpatients; primary care; population screening. This review therefore stratifies all analyses based on the following three populations:

1. People in population or community samples, regardless of whether they perceive a problem with their memory (screening);
2. People presenting to primary care practitioners with subjective memory problems that have not been previously assessed;
3. People referred to a secondary care outpatient clinic for the specialist assessment of memory difficulties.

The severity (stage) of dementia at diagnosis will influence the diagnostic accuracy of the MoCA. People with more advanced

disease will often score lower on the MoCA than those who are in the early stages. Therefore applying a low cut-point (or threshold indicative of normal cognition) will often result in the test being more specific (and might therefore be more applicable if the test is being used in a diagnostic work-up), and applying a high cut-point will often result in the test being more sensitive for identifying any possible cases of dementia (for instance if the MoCA is used for screening).

In the UK, people usually first present to their general practitioner (Figure 1). One or more brief cognitive tests (including the index test) may be administered, and might result in a referral to a memory clinic for specialist diagnosis (Boustani 2003; Alzheimer's Society 2013; Cordell 2013). However, many people with dementia may not present until later in the disorder and may follow a different pathway to diagnosis, e.g. referral to a community mental health team for individuals with complex problems otherwise unable to attend a memory clinic. Others may be identified during an assessment for an unrelated physical illness, e.g. during an outpatient appointment or an inpatient hospital admission.

In general, the role of non-specialist community services in dementia diagnosis is to recognise possible dementia and to refer on

to appropriate care providers, though this may vary geographically (Greening 2009; Greaves 2010). Some community settings have a higher prevalence of dementia than others. For example, the pre-test probability of prevalent dementia among residents in care homes is much higher than in the general population (Matthews 2002; Plassman 2007). This has led some to suggest routine cognitive assessment for every person resident in, or admitted to, a care home (Alzheimer's Society 2013). Through such an active case-finding strategy, a dementia diagnosis might be made outside the usual pathway.

Diagnostic assessment pathways vary across different countries, and diagnoses may be made by a variety of healthcare professionals including general practitioners, neurologists, psychiatrists, and geriatricians; thus we described the target populations rather than the exact setting in order to facilitate generalisability of the results. Common practice has been to use the recommended 25/26 cut-point (above 25 being considered normal) across the range of populations and settings (Nasreddine 2005). This review explicitly sought to assess the validity of this recommendation by stratifying the analyses by target population.

How might the index test improve diagnoses, treatments and patient outcomes?

The MoCA may help identify people requiring specialist assessment and treatment for dementia. Some symptomatic treatments and cognitive-behavioural interventions are available for people with dementia (Birks 2006; McShane 2006; Bahar-Fuchs 2013). Diagnosis of dementia can reduce uncertainty for individuals, their families and potential carers, facilitate access to appropriate services, and encourage planning for the future. However there is currently no curative treatment for dementia, which is a progressive disorder that eventually leads to death, so getting the diagnosis right (and reducing the risk of false positives) is important. Being wrongly tested as positive carries risk of significant costs and harm in the form of further unnecessary investigations or psychological distress, or both.

Outcomes for people with dementia in secondary care general hospital settings, including survival, length of stay and discharge to institutional care, are poor (RCPsych 2005; Sampson 2009; Zekry 2009). Accurate diagnosis may have specific benefits in addressing these adverse outcomes, in addition to facilitating access to the most suitable care and the use of non-pharmacological methods to manage behavioural and psychological symptoms of dementia. Diagnostic test accuracy of the MoCA or any test of cognitive function might be expected to differ in the hospital population due to the high prevalence of comorbid physical conditions that could mimic dementia and adversely affect performance on the test (hence the need to present the findings in different populations).

Rationale

The public health burden of cognitive and functional impairment due to dementia is a matter of growing concern. With the changing age structure of populations in both high- and low-income countries, overall societal costs from dementia are increasing (Wimo 2010). At the population level, this increased prevalence has major implications for service provision and planning, given that the condition leads to progressive functional dependence over several years. In the UK, it is estimated that the annual expenditure on dementia care is £26 billion (Alzheimer's Society 2014). Accurate and early diagnosis is crucial for planning care. In addition, accurate diagnosis is critical if participants for adequately powered clinical trials are to be identified.

Although dementia screening itself is not recommended by the United States Preventive Services Task Force (Boustani 2003) or the UK National Screening Committee, there appears to be a drift towards the opportunistic testing of older primary care attenders who have presented for reasons other than a memory complaint (Brunet 2012). The UK government has incentivised screening for dementia on acute admission to secondary care service, and also encourages the identification of dementia in people in primary care settings (Dementia CQUIN 2012). In the USA, the Patient Protection and Affordable Care Act (2010) added an Annual Wellness Visit, which includes a mandatory assessment of cognitive impairment (Cordell 2013). Therefore there is considerable value in determining the strength of evidence to support the use of the MoCA for identifying people with dementia.

OBJECTIVES

- To determine the diagnostic accuracy of the Montreal Cognitive Assessment (MoCA) at various thresholds for dementia and its subtypes, against a concurrently applied reference standard.

Secondary objectives

- To highlight the quality and quantity of research evidence available about the accuracy of the index test in the target populations;
- To investigate the heterogeneity of test accuracy in the included studies;
- To identify gaps in the evidence and where further research is required.

METHODS

Criteria for considering studies for this review

Types of studies

Criteria for including studies in this review were based on the generic protocol (Davis 2013a). We considered cross-sectional studies where participants received the index test and reference standard diagnostic assessment within three months (Davis 2013b). Participants needed to be derived from the same sample. We excluded case-control studies due to the high risk of spectrum bias. We did not include longitudinal studies (or related nested case-control studies), nor did we consider post-mortem verification of neuropathological diagnoses, as these designs are best evaluated through delayed-verification reviews (Davis 2013a).

Participants

We included all participants in whom the cross-sectional association between MoCA score and dementia ascertainment were described in a) general populations, regardless of health or residential status, b) primary care populations, c) memory clinic and other secondary care populations. The aim of the review was to establish the diagnostic accuracy of MoCA in unselected samples within these three target populations. Therefore, we excluded studies where participants were selected on the basis of a specific disease (for example Parkinson's disease, stroke patients, early-onset dementia, mild cognitive impairment). The diagnostic test accuracy of MoCA in these specific groups can be separately reviewed. We also excluded studies of participants with a secondary cause for cognitive impairment, e.g. current, or history of, alcohol/drug abuse, central nervous system trauma (e.g. subdural haematoma), tumour or infection.

Index tests

Any form of the Montreal Cognitive Assessment (details in Background) was included. We expected to find the recommended threshold of 25/26 to be used in studies to differentiate normal (26 and above) from impaired cognition (less than 26), however we also included studies using other thresholds.

Target conditions

Dementia, and any common dementia subtype including Alzheimer's disease, vascular dementia, Lewy body dementia and frontotemporal dementia.

Reference standards

We included studies that used a reference standard for all-cause dementia or any standardised definition of subtype as set out in the generic protocol (Davis 2013a). A number of clinical reference standards exist for dementia, e.g. International Classification of Diseases-10 (World Health Organization 1992), Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-

IV (American Psychiatric Association 1994) and the Clinical Dementia Rating scale (Morris 1993). There are also several reference standards for dementia sub-types, including the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for Alzheimer's disease dementia (McKhann 1984), the National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria for vascular dementia (Román 1993), Lewy body dementia (McKeith 2005), and frontotemporal dementia (Lund 1994). More recent studies might also have used new criteria using biomarkers to support a diagnostic classification (National Institute on Aging-Alzheimer's Association (NIA-AA) criteria (Jack 2012), DSM-5 (American Psychiatric Association 2013). This was a potential source of heterogeneity, which we could have explored quantitatively had a sufficient number of studies been identified (see below 'Investigations of heterogeneity').

We did not consider studies that applied a neuropathological diagnosis as this standard can only be applied to a diagnosis that is verified post-mortem; this could however be examined in a diagnostic test accuracy review of delayed verification studies.

Search methods for identification of studies

Electronic searches

We searched ALOIS (Cochrane Dementia and Cognitive Improvement Group specialized register of diagnostic and intervention studies), MEDLINE (OvidSP) (1950 to August 2012), EMBASE (OvidSP) (1974 to 01 August 2012), PsycINFO (OvidSP) (1806 to July week 4 2012), BIOSIS Previews (Web of Knowledge) (1945 to July 2012), Science Citation Index (ISI Web of Knowledge) (1945 to July 2012), and LILACS (Bireme). See Appendix 1 for details of the sources searched, the search strategies used, and the number of hits retrieved.

There was no attempt to restrict the search to studies based on sampling frame or setting as terms describing these types of studies are not standardised or consistently applied and would have been likely to reduce the sensitivity of the search. We did not use search filters (collections of terms aimed at reducing the number needed to screen) as an overall limiter because those published have not proved sensitive enough (Whiting 2011b; Beynon 2013). We did not apply any language restriction to the searches.

A single researcher with extensive experience of systematic reviews conducted the searches.

Searching other resources

Reference lists of all relevant papers were checked for possible additional studies. We also searched:

- MEDION database (Meta-analyses van Diagnostisch Onderzoek: www.york.ac.uk/inst/crd/crddatabases.html);
- DARE (Database of Abstracts of Reviews of Effects: (<http://www.crd.york.ac.uk/CRDWeb/>);
- HTA Database (Health Technology Assessments Database, via *The Cochrane Library*);
- ARIF database (Aggressive Research Intelligence Facility: www.arif.bham.ac.uk);
- C-EBLM (International Federation of Clinical Chemistry and Laboratory Medicine Committee for Evidence-based Laboratory Medicine).

We used the ‘related articles’ feature in PubMed and tracked key studies in the Science Citation Index and Scopus to identify any further relevant studies. We searched grey literature through the Web of Science Core Collection, including Science Citation Index and Conference Proceedings Citation Index (Thomson Reuters

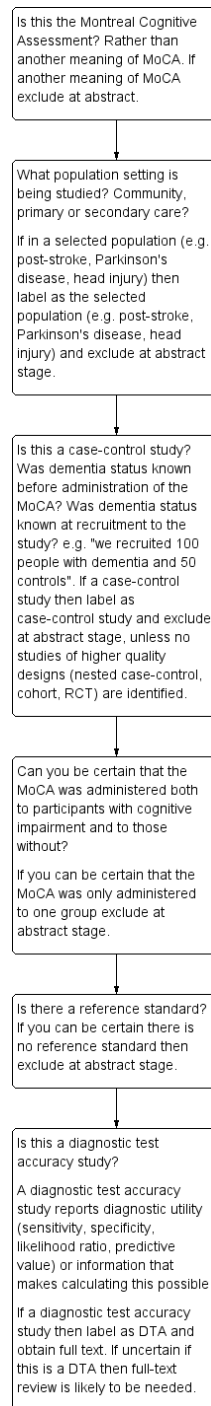
Web of Science) and attempted to access theses or PhD abstracts from institutions known to be involved in prospective dementia studies. We attempted to contact researchers involved in studies with possibly relevant but unpublished data. We did not perform handsearching as there is little evidence to support this at present (Glanville 2012).

Data collection and analysis

Selection of studies

Figure 2 shows a flowchart for inclusion of studies in the review. Two authors reviewed the title, abstract, and full text where needed, of all retrieved citations and considered whether the study should be included. Inclusion criteria were:

Figure 2. Flow diagram for inclusion of articles in review



- Participants as described above;
- Reference standard as described above;
- MoCA is used as an index test (may have been included as one of several tests).

Exclusion criteria were:

- Participants wholly drawn from selected populations (e.g. post-head injury, stroke patients, people with Parkinson's disease only, nursing home residents);
- If diagnostic accuracy could not be calculated because index test and reference standard are not applied to both people with and without dementia.

Data extraction and management

Two authors worked independently to extract data on study characteristics to a pro forma, which included data for the assessment of quality and for the investigation of heterogeneity (details are given in [Appendix 2](#)). We piloted the pro forma against the first three studies and refined accordingly. The results were dichotomised where necessary and cross-tabulated in two-by-two tables of index test result (positive or negative) against target disorder (positive or negative) directly into Review Manager 5 (RevMan) tables.

Assessment of methodological quality

Two authors independently assessed, discussed and reached a consensus on the methodological quality of each study using the QUADAS-2 tool ([Whiting 2011a](#)) as recommended by Cochrane. Operational definitions describing the use of QUADAS-2 for MoCA are detailed in [Appendix 3](#).

Statistical analysis and data synthesis

The target condition comprised two categories: (1) all-cause dementia, and (2) dementia subtype. For all included studies, we used the data in the two-by-two tables (showing the binary test results cross-classified with the binary reference standard) to calculate the sensitivities and specificities, with their 95% confidence intervals. We presented individual study results graphically by plotting estimates of sensitivities and specificities in both a forest plot and in receiver operating characteristic (ROC) space. Our systematic assessment of study quality used QUADAS-2 to help determine the overall risk of bias for each study and therefore the strength of evidence from the study.

For studies that reported more than one threshold, we extracted diagnostic accuracy data for each threshold and presented these for each study. We intended to use only the threshold of 25/26 (over 25 indicating normal cognition) to estimate the summary

statistics for diagnostic accuracy in each of the three populations, in order to avoid studies that contributed data from more than one threshold being included in the calculation of a summary statistic on more than one occasion (in the same population).

We intended to perform a meta-analysis on pairs of sensitivity and specificity if it was appropriate to pool the data. Had the majority of the studies in each of the three populations reported results with consistent thresholds, a bivariate random-effects approach based on pairs of sensitivity and specificity would have been appropriate ([Reitsma 2005](#)).

Investigations of heterogeneity

There are many potential sources of heterogeneity for this review (see [Davis 2013a](#) for a more extensive account). We intended to investigate heterogeneity due to the age distribution of the sample, MoCA threshold score used to diagnose possible all-cause dementia or its subtypes, the reference standard used and the severity of the target disorder. Heterogeneity due to disease severity was addressed in part by the QUADAS assessments of spectrum bias. Differences in diagnostic test accuracy were expected a priori in the identified target populations described above and are thus presented separately.

Sensitivity analyses

Had a quantitative synthesis been possible, our intention had been to perform sensitivity analyses to determine the effect of excluding studies that were deemed to be at high risk of bias according to the QUADAS-2 checklist, for example studies where blinding procedures have not been clearly described.

Assessment of reporting bias

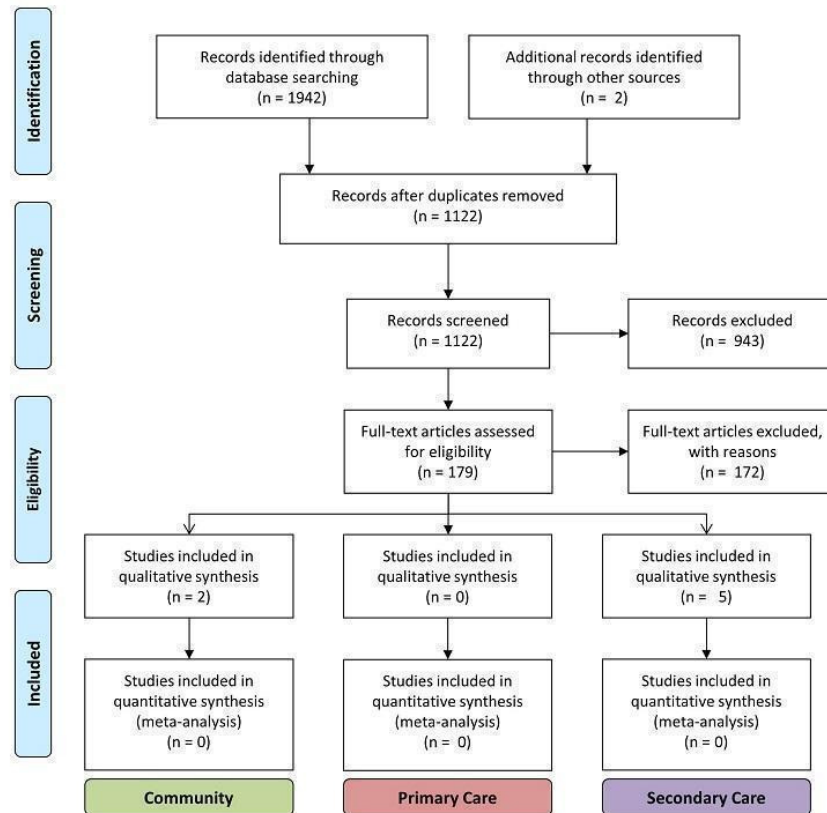
Quantitative methods for exploring reporting bias are not well established for studies of DTA. Specifically, funnel plots of the diagnostic odds ratio (DOR) versus the standard error of this estimate were not considered.

RESULTS

Results of the search

The PRISMA flow diagram detailing the search, assessment for eligibility and inclusion is detailed in [Figure 3](#). 172/179 studies were excluded for the following reasons:

Figure 3. Flow diagram showing study selection, by setting. Reasons for exclusion: Not a DTA study 12; Mild cognitive impairment given as outcome 9; Not a MoCA study 6; Selected secondary care 84 (vascular dementia 31; Huntington's dementia 10; diabetes cohort 1; Parkinson's disease 41; nursing home cohort 1); Delayed-verification 1; No dementia outcome 29; Reference standard not explicit 3; No normal controls 2; Duplicated data (abstract) 14; Case control design 11; Insufficient DTA data reported 1. Excluded studies that were considered in detail (including contact with authors) are fully referenced (n = 14).



- Not a diagnostic test accuracy study 12;
- Reference standard of mild cognitive impairment, not dementia 9;
 - MoCA not index test 6;
 - Population selected either for specific disease or place of residence 84 (cerebrovascular disease 31; Huntington's disease 10; diabetes cohort 1; Parkinson's disease 41; nursing home cohort 1);
 - Delayed-verification design 1;
 - No dementia outcome 29;
 - Reference standard not explicit 3;
 - No non-diseased group 2;
 - Duplicated data (abstract) 14;
 - Case control design 11;
 - Insufficient DTA data reported 1.

Twenty-one studies were initially thought to be eligible for inclusion; however, on closer inspection by two of the reviewers (DD, SC) and correspondence with most of the study authors, only 7 of the 21 studies met the inclusion criteria: five in secondary care (two in geriatric medicine outpatients (Lee 2008; Cecato 2011); and three in dedicated memory clinics (Smith 2007; Chen 2011; Larner 2012)), none in primary care, and two in population-derived samples (Kasai 2011; Lu 2011). The characteristics of the 14 excluded studies are presented in the [Characteristics of excluded studies](#) table.

Methodological quality of included studies

The methodological quality in each of the domains was often difficult to assess as the required information was not clearly stated

in the published reports. In our assessments, we contacted 14 authors of the 21 initially included studies for further information on methodological quality; all responded and seven studies were included in the final review.

The QUADAS-2 scores for each domain are shown in Figure 4 and Figure 5. For *Patient Selection*, only studies that avoided a case-control design were included, which reduced the risk of selection bias. One of the seven studies selected the majority of the sample consecutively from the population attending the hospital-based

clinic (but included a small number of volunteers) and therefore this study was included, but with 'unclear' risk of selection bias (Lee 2008). Three studies clearly reported that they had selected a random or consecutive sample of participants (Chen 2011; Lu 2011; Larner 2012). Only one study commented on exclusions from the sample (Cecato 2011); the remainder were thus rated as 'unclear' for this question. Three of the studies were rated as having low risk of concerns about applicability (Chen 2011; Lu 2011; Larner 2012).

Figure 4. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies

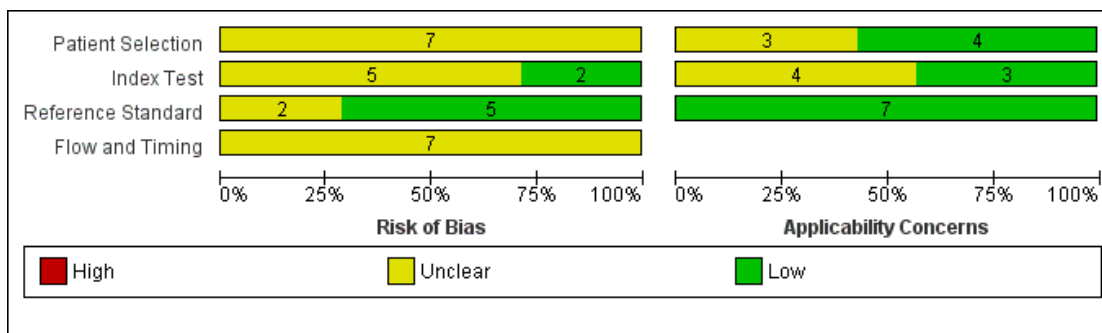





Figure 5. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Cecato 2011	?	?	+	?	+	?	+
Chen 2011	?	?	?	?	+	?	+
Kasai 2011	?	?	+	?	?	?	+
Larner 2012	?	?	+	?	+	+	+
Lee 2008	?	+	+	?	?	+	+
Lu 2011	?	+	+	?	+	+	+
Smith 2007	?	?	?	?	?	?	+

 High	 Unclear	 Low
---	--	--

In the *Index Test* domain only two studies were judged as being at low risk of bias (Lee 2008; Lu 2011). In five studies the reporting on test administration, masking and pre-specification of thresholds was absent or unclear and so the studies were judged as being of unclear risk of bias. Three studies were considered to have low risk of concerns about applicability (Lee 2008; Lu 2011; Larner 2012).

In the *Reference Standard* domain all studies used a recognised reference standard that was likely to correctly classify the condition, but only four of the seven studies reported use of masking (Lee 2008; Cecato 2011; Lu 2011; Larner 2012); and only two

reported how the reference standard was operationalised and applied (Lu 2011; Larner 2012). These two studies (Larner 2012 and Lu 2011) were assessed as being of low risk of concern about applicability.

All seven studies were judged as unclear for flow and timing due to absence of relevant information: only three studies reported the time interval between index test and reference standard (Chen 2011; Lu 2011; Larner 2012); none of the studies provided a flow diagram; it was not always clear if the reference standard was consistently applied across the whole sample; and no article

reported if all participants were included in the final analysis.

Findings

The main aim of the review was to present the diagnostic accuracy of the MoCA in three different populations.

By setting

a) General population (participants selected regardless of whether they perceive a problem with their memory)

Two of the included studies were carried out in general population cohorts in Japan (Kasai 2011), and China (Lu 2011), in which the prevalence of dementia was between 10% and 5% respectively. Only one of these reported the diagnostic accuracy of the MoCA at the recommended threshold of 25/26; sensitivity was 0.98 and specificity was 0.52 (Lu 2011). Kasai 2011 used a lower threshold of 20/21 and reported sensitivity of 0.77 and specificity of 0.57.

b) Primary care population (people presenting to primary care practitioners with subjective memory problems that have not been previously assessed)

No studies were carried out in this population.

c) Secondary care outpatient population (people referred to a secondary care outpatient clinic for the specialist assessment of memory difficulties)

Five studies were carried out in outpatient clinic populations: three in specialist memory clinics in Singapore (Chen 2011) and UK (Larner 2012; Smith 2007), and two in geriatric outpatient clinics in Brazil (Cecato 2011) and Korea (Lee 2008). Two of the three studies based in memory clinics had a higher prevalence of dementia: 54% (Chen 2011) and 48% (Smith 2007), whereas Larner 2012 reported a lower prevalence of 24% in a population with a lower median age of 61; geriatric clinic populations had slightly lower dementia prevalence: 28% (Cecato 2011) and 22% (Lee 2008). Three out of the five secondary care-based studies reported diagnostic accuracy at a cut-point of 25/26: sensitivity ranged from 0.94 to 1.00 and specificity from 0.50 to 0.60; the other two used lower thresholds: Cecato 2011 reported sensitivity/specificity of 0.96/0.88 at cut-point 24/25 and Chen 2011 reported sensitivity/specificity of 0.94/0.66 at cut-point 18/19. It is not clear why sensitivity was high and specificity was low at the low threshold reported in Chen 2011, as all other studies reported lower sensitivity and higher specificity at the lower cut-points; it may be due to the low educational level (mean = 6 years) of this sample compared to the other secondary care-based studies or it may have been due to the use of a previously unvalidated Singaporean-MoCA scale.

All settings

The characteristics of all included studies are summarised in [Summary of findings](#). Studies ranged in size from 53 to 8411 participants (total n = 9422), most including people aged between 70 and 75 years. Two studies sampled from general populations in

Japan and in China, and five sampled populations attending outpatient clinics in Singapore, Brazil, Korea and UK. The prevalence of dementia in the included studies varied with setting; studies based in memory clinics had the highest prevalence of dementia (24% to 54%) whereas geriatric clinics had slightly lower prevalence (22% to 28%) and studies based in the community had a prevalence of dementia of 5% and 10%.

The target condition: dementia, subtypes, with or without MCI

The target condition was all-cause dementia in five studies and Alzheimer's disease dementia in two studies. The reference standards used to diagnose all-cause dementia were DSM-IV (Chen 2011; Lu 2011; Larner 2012), ICD-10 (Smith 2007) and CDR (Kasai 2011). Alzheimer's disease dementia was diagnosed using NINCDS-ADRDA (Cecato 2011) and DSM-IV (Lee 2008). Studies were not consistent in how they reported their findings, specifically in respect of whether MCI was included as part of the target condition dementia or as part of non-dementia. Three studies reported the diagnostic test accuracy of MoCA in distinguishing dementia from non-dementia (including MCI), which is our preferred classification (Smith 2007; Cecato 2011; Chen 2011). Others reported the diagnostic test accuracy for cognitive impairment (that is, dementia + MCI) versus normal (Kasai 2011; Lu 2011; Larner 2012), and one study excluded participants with MCI from the analysis altogether, which might increase spectrum bias (Lee 2008).

The effect of threshold

The MoCA is scored out of a possible 30 points, and the recommended threshold is 25/26 (26 and over indicating normal cognitive function). Three of the five studies in secondary care setting reported diagnostic accuracy at a cut-point of 25/26 (Smith 2007; Lee 2008; Larner 2012), as did one study in the general population (Lu 2011). The other studies reported diagnostic accuracy at alternative thresholds ranging from 18/19 to 26/27. Lee 2008 reported sensitivity and specificity for a range of thresholds from 18/19 to 26/27 in a sample ascertained from a geriatric outpatient population. As might be expected, the sensitivity increased and specificity decreased when the MoCA cut-point was higher; for example sensitivity/specificity was 0.86/0.97 at cut-point 18/19 but changed to 1.00/0.34 at the cut-point of 26/27 on the MoCA. [Summary of findings](#), [Figure 6](#) and [Figure 7](#) summarise the studies and their reported or extracted sensitivity and specificity data. Studies based in secondary care settings reported sensitivity ranging from 0.63 at threshold of 19/20 (Larner 2012) to 1.00 at threshold of 26/27 (Lee 2008), and specificity ranging from 0.34 at 26/27 (Lee 2008) to 0.97 at 18/19 (Lee 2008).

Figure 6. Forest plot of MoCA in secondary care populations at different thresholds (above threshold considered normal). Thresholds are for reference thresholds (unadjusted).

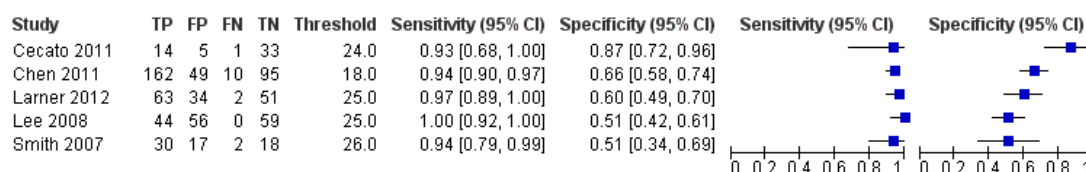
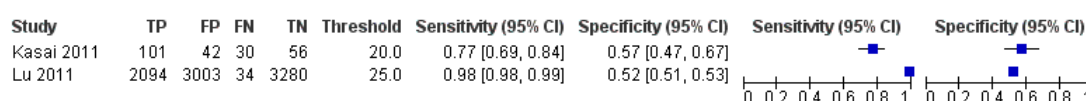


Figure 7. Forest plot of MoCA in community-based populations at different thresholds (above threshold considered normal). Thresholds are for reference thresholds (unadjusted).



Four studies reported at the recommended cut-point of 26 and over indicating normal cognition: two in UK memory clinics (Larner 2012; Smith 2007); one in a Korean geriatric clinic (Lee 2008); and one in a Chinese population (Lu 2011). Despite these different settings, the reported sensitivity (0.94 to 1.00) and specificity (0.50 to 0.60) was similar.

Other aspects of study design

Each of the seven included studies applied the MoCA contemporaneously with a reference standard (within three months) in a cohort, but there was wide variation in other elements of study design (see Summary of findings). The seven study samples were selected from six different countries (Japan, China, Singapore, Ko-

rea, Brazil and UK) and three different populations (general population, geriatric clinic and specialist memory clinic); they ranged in sample size (53 to 8411), in reference standard used (DSM-IV, ICD-10, CDR and NINCDS-ADRDA) and in MoCA threshold (from 18/19 to 26/27). Overall, these differences resulted in considerable heterogeneity, even within similar populations. The handling of MCI with respect to the primary comparisons was very inconsistent. Either the reported outcome was dementia alone, or it was dementia and MCI. Further contact with authors did not yield disaggregated data, making very few studies directly comparable. Because of these and other differences in study design, participant sampling, reference standards and reported thresholds, we judged that a meta-analysis was not appropriate.

Summary of findings

Patient population	Participants who met criteria for inclusion in three populations: <ol style="list-style-type: none"> 1. People in the general population, regardless of whether they perceive a problem with their memory (screening) 2. People presenting to primary care practitioners with subjective memory problems that have not been previously assessed 3. People referred to a secondary care outpatient clinic for the specialist assessment of memory difficulties 						
Prior testing	Prior testing in secondary care populations likely to include a cognitive screening tool and clinical judgement						
Index test	MoCA including versions in different languages						
Reference standard	DSM or ICD criteria for dementia; NINCDS-ADRDA or for Alzheimer's disease dementia; McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; and NINDS-AIREN criteria for vascular dementia						
Target condition	All cause dementia; Alzheimer's disease dementia.						
Included studies	7 studies (9422 participants)						
Quality concerns	Most studies were at low or unclear risk of bias for patient selection, index test, reference standard, flow and timing; unclear risk was usually due to lack of transparent reporting of methods in studies Applicability: most studies were considered to have reasonably high applicability for patient populations, index test and reference standards						
Heterogeneity	Considerable heterogeneity due to varied settings, countries, populations, reference standards and versions of index tests						
INCLUDED STUDIES							
Author, year	Chen 2011	Cecato 2011	Larner 2012	Lee 2008	Smith 2007	Lu 2011	Kasai 2011
Population and setting	Memory clinic Singapore	Geriatric clinic Brazil	Memory clinic UK	Geriatric clinic Korea	Memory clinic UK	Population-based China	Population-based Japan
Sample size	316	53	150	196 (37 MCI)	67	8411	229
Age	73 (SD ± 10)	76 (SD ± 7.4)	61 (range 20 to 87)	70 (SD ± 6.3)	74 (SD ± 10)	73 (SD ± 0.9)	≥ 75 years

Education	mean 6 years	66% > 7 years	not reported	normal 8.0 years MCI 8.3 years dementia 7.9 years	mean 12.1 years	normal 5.8 years dementia 2.5 years	normal 9.5 years MCI 8.6 years dementia 8.0 years
Index test	MoCA - Singaporean	MoCA - Brazilian	MoCA - original	MoCA - Korean	MoCA - original	MoCA - Beijing	MoCA - Japanese
Target condition and prevalence	All cause dementia 54% (subtypes not specified)	Alzheimer's disease 28%	All cause dementia 24% (subtypes not specified)	Alzheimer's disease 22%	All-cause dementia 48% (AD 27%, VD 19%, PDD 2%)	All-cause dementia 5% (AD 3.5%, VD 1%, Other 0.5%)	All-cause dementia 10% (subtypes not specified)
Reference standard	DSM-IV	DSM-IV NINDS-ADRDA	DSM-IV	DSM-IV and CDR	ICD-10	DSM-IV NINDS-ADRDA NINDS-AIREN	CDR
Comparison groups	Dementia vs (normal + MCI)	Dementia vs (normal + MCI)	(Dementia + MCI) vs normal	AD vs normal (37 MCI excluded from analysis)	Dementia vs (normal + MCI)	(Dementia + MCI) vs normal	(Dementia + MCI) vs normal
Sensitivity/specificity by threshold*							
17/18							
18/19	0.94/0.66			0.86/0.97			
19/20			0.63/0.95		0.93/0.97		
20/21				0.93/0.94		0.77/0.57	
21/22				0.98/0.90			
22/23				0.98/0.84			
23/24				0.98/0.76			
24/25	0.96/0.88		0.98/0.61				

25/26	Data not provided by author	Data not provided by author	0.97/0.60	1.00/0.50	0.94/0.50	0.98/0.52	Data not provided by author
26/27				1.00/0.34	0.97/0.35		
Conclusions	<p>Only seven studies were found that did not use case control design: two in general (unscreened) populations, none in primary care populations and five in secondary care populations. Five different language versions of MoCA were used in six countries. The target condition was all-cause dementia in five studies and Alzheimer's disease dementia in two studies. MCI was analysed as a non-disease state in three studies, included with dementia as a disease state in three studies and excluded from the analysis altogether in one study</p> <p>Due to the considerable heterogeneity a meta-analysis was not performed</p> <p>Four studies used the recommended threshold of 25/26 in three secondary care populations and one population-based study. In both target populations the sensitivity was high (range 0.94 to 1.00) and the specificity was low (range 0.50 to 0.60). This is likely to be due to the fact the MoCA was developed to detect MCI and thus the recommended cut-point is too high for optimal detection of dementia (minimising both false positives and false negatives)</p>						
Implications	<p>Further research is required to examine the optimal diagnostic accuracy of MoCA in the detection of dementia in different target populations. We recommend that the STARDdem standards are used to guide adequate reporting of DTA study findings</p>						

* where multiple results are reported by a study, these are from the same participants in one single sample

DISCUSSION

Summary of main results

Our search to identify diagnostic test accuracy studies of the MoCA in a range of settings retrieved predominantly case-control studies, which were excluded from this review because they are at high risk of (selection) bias, leaving just seven included studies. Methodological differences precluded a meta-analysis of these seven studies. Most of the included studies were of relatively small sample size, four had fewer than 200 participants and only one study had more than 350 participants. In the four studies conducted at the recommended threshold of 26 or above indicating normal cognition, the MoCA had high sensitivity of 0.94 and above but low specificity of 0.60 and below (Smith 2007; Lee 2008; Lu 2011; Lerner 2012).

Strengths and weaknesses of the review

We conducted our review as specified in the published protocol to facilitate robust and rigorous decision-making at each stage of the review process (Davis 2013b). The search was extensive and included multiple databases and grey literature sources. To maximise sensitivity, the search approach used a single-concept search across multiple databases combined with a search of ALOIS, the Cochrane Dementia and Cognitive Improvement Group's burgeoning register of diagnostic test accuracy studies. For this review, the majority of studies would be identified by terms related to this index test, the terminology for which is reasonably standardised. In addition, to capture the harder-to-locate studies where the index test is not referred to specifically in the parts of the electronic record available for search retrieval, searching the Group's register, a unique source which is populated regularly using a more complex multi-concept 'generic' search in major biomedical databases, should ensure that studies are also retrieved where this specific index test is included but unnamed, for example as part of a report on a range of neuropsychological tests. To enhance methodological rigour of the review process, the citation screening, quality assessment and data extraction were performed by at least two assessors. We approached the original authors for further information when needed to resolve uncertainty arising from insufficient reporting. Had the included studies enabled us to draw more definitive conclusions, the review would likely be more useful to clinicians, commissioners of health services and the public, as the MoCA is now commonly used for cognitive testing in place of the previously popular Mini-Mental State Examination or MMSE (Folstein 1975).

Applicability of findings to the review question

In a secondary care setting, the MoCA appears to achieve a high sensitivity (94%, 97%) with modest specificity (50%, 60%) at the threshold of 26 and above out of 30 indicating normal cognition (Smith 2007; Lerner 2012). In a general population sample in China, also using a threshold of 26 and above indicating normal cognitive function, sensitivity remains high (98%), with similar specificity (52%) (Lu 2011). Furthermore, comparing the findings for all studies reporting at the threshold of 25/26, neither the definition of the target condition nor reference standard used made much difference to the reported sensitivities and specificities at this threshold; all of the studies reported high sensitivity and low specificity at this threshold (Smith 2007; Lee 2008; Lu 2011; Lerner 2012). We would expect the diagnostic accuracy of the MoCA to vary more across population settings and the consistently high reported sensitivity may reflect a ceiling effect of the MoCA at the recommended threshold of 25/26, suggesting that this threshold is too high. The original instructions for use of the MoCA recommended a threshold of 25/26 in order to identify *any* cognitive impairment (i.e. to have high sensitivity for MCI or dementia) (Nasreddine 2005); but this is not necessarily the *optimum* threshold for diagnosing dementia, which may be several points lower.

The optimum threshold of a test for diagnosing dementia is one that maximises both sensitivity and specificity thus reducing the number of false positives whilst also minimising the number of false negatives. For example, applying the Lu 2011 results to a hypothetical general population of 1000 people for screening, in whom the prevalence of dementia would be around 6.5% in the UK (Matthews 2013), we would predict 65 people would have dementia. At the current threshold of 25/26, one of the 65 people with dementia would score 26 or more on the MoCA and so the diagnosis would be 'missed', but 449 people *without* dementia would score below 26, mandating further work-up and (potentially invasive) tests, to exclude dementia. Approximately 7 (449/64) people would need further evaluation to identify one case of dementia and so there would be a significant potential that some people might be harmed unnecessarily (either emotionally, physically, or financially). In an alternative setting, the prevalence of dementia is higher, for example in a memory clinic where the prevalence of all-cause dementia might be as high as 50%. Applying the Smith 2007 results to a hypothetical memory clinic of 1000 people we would predict that, at a threshold of 25/26, of the 500 people with dementia, on average 30 would be incorrectly diagnosed as having normal cognition (score 26 or above on the MoCA, false negatives) but that on average 250 people of 500 without dementia would receive a false positive diagnosis of dementia (score below 26). Thus in both settings the number of false positives might be judged as too high to be ethical or cost-effective.

It is worth noting that the original study was a case-control design which calculated the normal data from a sample of 90 healthy elderly Canadian controls and compared with MCI and dementia

cases ascertained from a tertiary care memory clinic (Nasreddine 2005). This is likely to have distorted the findings due to spectrum bias (only the extremes of the spectrum of cognitive function were included) and may have resulted in the original recommended threshold for cognitive impairment being too high. The studies in our review have shown that many 'normal' people (i.e. without dementia or MCI) score below the recommended cut-point. Some of the included studies attempted to explore the optimal cut-points for detecting cognitive impairment. In by far the largest study in this review (n = 8411), Lu 2011 adjusted the recommended thresholds to find the optimal cut-point for detecting cognitive impairment (MCI or dementia) in individuals with no formal education, with 1 to 6 years of education and with less than 6 years of education. For individuals with no formal education the most appropriate MoCA cutoff was 13/14 (n = 2279, sensitivity 80.9% and specificity 83.2%); for individuals with 1 to 6 years of education, the most appropriate MoCA cutoff was 19/20 (n = 3085, sensitivity 83.8% and specificity 82.5%); and for individuals with 7 or more years of education, it was 24/25 (n = 3047, sensitivity 89.9% and specificity 81.5%). After applying the adjusted cut-points, the sensitivity of the MoCA was 83.8% for any cognitive impairments (MCI or dementia), but still remained high at 96.9% for dementia. This study was carried out in urban and rural China, in participants with very low educational levels, and would thus not necessarily apply to other populations. Further studies in different populations are required to identify the optimum thresholds of MoCA for diagnosing dementia.

AUTHORS' CONCLUSIONS

Implications for practice

The overall quality and quantity of information is insufficient to be able to make recommendations on the clinical utility of the MoCA for diagnosing dementia. In population settings where the prevalence of dementia is low, many people without dementia would score below the traditional cut-point of 26, and would require further testing. In secondary care settings a smaller number of people without dementia would score below 26, and those that did would expect further investigations by virtue of having been referred to a specialist clinic. Nonetheless the findings from both population and secondary care settings suggest that the optimal threshold for the MoCA should be lower than 25/26 for efficient use of resources. Overall we would recommend against approaches that use the MoCA in isolation, regardless of the expected prevalence of dementia in the clinical setting.

Implications for research

Further research is required to examine the optimum threshold of the MoCA for the diagnosis of dementia and its sub-types, and for

the thresholds to be tested in different healthcare populations and in different countries. Studies investigating the diagnostic test accuracy of the MoCA should aim to recruit a cohort of participants that reflects the target population (for example the general population, primary care population, outpatient population, hospital inpatients, with the ability to explore cross-cultural differences across various countries) and administer the index test and reference standard contemporaneously, to reflect the clinical situation. The prevalence of dementia, and alternative diagnoses in the population should be clearly reported, making reference to STARD or STARDdem (Noel-Storr 2014). Investigators might consider evaluating how the MoCA performs at different thresholds, across populations with varying prevalence of disease.

Delayed verification study designs (prospective cohort studies, nested case control studies embedded in cohort designs or both) might also be useful where the reference standard is applied prospectively after a period of at least one year, either with or without neuropathological confirmation of diagnosis after death, as described in Davis 2013a. These studies are likely to give more accurate findings for dementia diagnosis but are also more expensive and time-consuming.

Researchers should also consider evaluating and reporting the incremental or added value of the MoCA in a diagnostic work-up, so that clinicians can understand more about how the use of the MoCA changes patient-relevant outcomes (for example, harms of unnecessary testing against benefit of earlier diagnosis), and how the burden of the MoCA for individuals and clinicians weighs against its potential benefits.

Further reviews and meta-analyses can be carried out when there are sufficient studies in the defined populations.

ACKNOWLEDGEMENTS

The following persons kindly provided additional information on individual studies: Julienne Aulwes, University of Hawaii; Jun Young Lee, Seoul National University College of Medicine; Jianping Jia, Beijing Capital Medical University; Juliana Cecato, Jundiai Medical School; Kenichi Meguro, Tohoku University; Christopher Chen, National University Singapore; Varuni de Silva, University of Colombo; Floor van Bergen, Jeroen Bosch Ziekenhuis; Jong Ling Fuh, Veterans General Hospital, Taipei; Nai Ching Chen, Chang Gung College of Medicine, Kaohsiung; Ana Costa, RWTH Aachen University; Yoshinori Fujiwara, Tokyo Metropolitan Institute of Gerontology; Cheryl Luis, Roskamp Institute, Florida; Clive Holmes, University of Southampton.

In addition, the following individuals helped with translations of articles: Emily Zhao, University of Cambridge; Rianne M. van der

Linde, University of Cambridge; Gul Nihal Ozdemir, Istanbul University.

REFERENCES

References to studies included in this review

Cecato 2011 *{published and unpublished data}*

J Cecato, JE Martinelli, I Aprahamian, M Yassuda. MoCA contributions to differential diagnosis among normal controls, mild cognitive impairment and Alzheimer's disease in Brazil. Conference: Alzheimer's Association International Conference, Paris France. Conference Start: 20110716 Conference End: 20110721. 2011.

Chen 2011 *{published and unpublished data}*

C Chen, YH Dong, R Merchant, S Collinson, E Ting, SL Quah, et al. The Montreal Cognitive Assessment (MoCA) is superior to the Mini-Mental State Examination (MMSE) in detecting patients with moderate cognitive impairment, no-dementia (CIND) and at high risk of dementia. Conference: Alzheimer's Association International Conference, Paris France. Conference Start: 20110716 Conference End: 20110721. 2011.

Kasai 2011 *{published and unpublished data}*

M Kasai, K Nakamura, Y Kato, M Nakai, M Nakatsuka, H Ishikawa, et al. A community-based study of the Montreal Cognitive Assessment (MoCA) in Japan: The Kurihara project. Conference: Alzheimer's Association International Conference, Paris France. Conference Start: 20110716 Conference End: 20110721. 2011.

Larner 2012 *{published data only}*

AJ Larner. Screening utility of the Montreal Cognitive Assessment (MoCA): in place of--or as well as--the MMSE?. *International Psychogeriatrics* 2012;**24**(3):391-6.

Lee 2008 *{published and unpublished data}*

JY Lee, Lee Dong Woo, SJ Cho, DL Na, Jeon Hong Jin, SK Kim, et al. Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology* 2008;**21**(2):104-10.

Lu 2011 *{published and unpublished data}*

J Lu, D Li, F Li, A Zhou, F Wang, X Zuo, et al. Montreal cognitive assessment in detecting cognitive impairment in Chinese elderly individuals: a population-based study. *Journal of Geriatric Psychiatry and Neurology* 2011;**24**(4):184-90.

Smith 2007 *{published and unpublished data}*

T Smith, N Gildeh, C Holmes. The Montreal Cognitive Assessment: validity and utility in a memory clinic setting. *Canadian Journal of Psychiatry* 2007;**52**(5):329-32.

References to studies excluded from this review

Aulwes 2012 *{published and unpublished data}*

JO Aulwes, S Nishimura, R Suzuki, D Goebert, B Lu, I Ahmed, et al. Comparison of the Saint Louis University

Mental Status examination (SLUMS), the Montreal Cognitive Assessment (MoCA), and the Mini-Mental State Examination (MMSE) for detecting mild neurocognitive disorder and dementia - A pilot study. Conference: AAGP Annual Meeting 2012 Washington, DC United States. Conference Start: 20120316 Conference End: 20120319. Conference Publication:.. 2012.

Chang 2012 *{published and unpublished data}*

YT Chang, CC Chang, HS Lin, CW Huang, WN Chang, CC Lui, et al. Montreal Cognitive Assessment in assessing clinical severity and white matter hyperintensity in Alzheimer's disease with normal control comparison. *Acta Neurologica Taiwanica* 2012;**21**(2):64-73.

Costa 2012 *{published and unpublished data}*

AS Costa, B Fimm, P Friesen, H Soundjock, C Rottschy, T Gross, et al. Alternate-form reliability of the Montreal Cognitive Assessment screening test in a clinical setting. *Dement Geriatr Cogn Disord* 2012;**33**:379-84.

Fujiwara 2010 *{published and unpublished data}*

Y Fujiwara, H Suzuki, M Yasunaga, M Sugiyama, M Ijuin, N Sakuma, et al. Brief screening tool for mild cognitive impairment in older Japanese: validation of the Japanese version of the Montreal Cognitive Assessment. *Geriatrics and Gerontology International* 2010;**10**(3):225-32.

Guo 2010 *{published data only}*

QH Guo, XY Cao, Y Zhou, QH Zhao, D Ding, Z Hong. Application study of quick cognitive screening test in identifying mild cognitive impairment. *Neuroscience Bulletin* 2010;**26**(1):47-54.

Kandiah 2011 *{published data only}*

N Kandiah, Y Jion, A Ng. Severity of cognitive impairment is correlated to levels of vitamin D. Conference: Alzheimer's Association International Conference, Paris France. Conference Start: 20110716 Conference End: 20110721. 2011.

Karunaratne 2011 *{published and unpublished data}*

S Karunaratne, R Hanwella, V de Silva. Validation of the Sinhala version of the Montreal Cognitive Assessment in screening for dementia. *The Ceylon Medical Journal* 2011;**56**(4):147-53.

Luis 2009 *{published and unpublished data}*

CA Luis, AP Keegan, M Mullan. Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *International Journal of Geriatric Psychiatry* 2009;**24**(2):197-201.

Magierska 2012 *{published data only}*

J Magierska, R Magierski, W Fendler, I Kloszewska, TM Sobow. Clinical application of the Polish adaptation of the Montreal Cognitive Assessment (MoCA) test in screening

- for cognitive impairment. *Neurologia i Neurochirurgia Polska* 2012;**46**(2):130–9.
- Nasreddine 2005** *{published data only}*
ZS Nasreddine, NA Phillips, V Bedirian, S Charbonneau, V Whitehead, I Collin, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* 2005; **53**(4):695–9.
- Selekler 2009** *{published data only}*
K Selekler, B Cangoz, S Uluc. Power of discrimination of Montreal Cognitive Assessment (MOCA) scale in Turkish patients with mild cognitive impairment and Alzheimer's disease. *Türk Geriatri Dergisi* 2010;**13**(3):166–71.
- Thissen 2010** *{published and unpublished data}*
AJ Thissen, F van Bergen, JF de Jonghe, RP Kessels, PL Dautzenberg. Applicability and validity of the Dutch version of the Montreal Cognitive Assessment (MoCA-d) in diagnosing MCI. *Tijdschrift voor Gerontologie en Geriatrie* 2010;**41**(6):231–40.
- Tsai 2012** *{published and unpublished data}*
C-F Tsai, W-J Lee, S-J Wang, B-C Shia, Z Nasreddine, J-L Fuh. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: Validation of the Taiwanese version of the MoCA and an item response theory analysis. *International Psychogeriatrics* 2012;**24**(4):651–8.
- Zhang 2008** *{published data only}*
LX Zhang, XQ Liu. Determination of the cut-off point of the Chinese version of the Montreal Cognitive Assessment among Chinese elderly in Guangzhou. *Zhongguo Xin li wei sheng Zazhi* 2008;**23**:123–125.
- Additional references**
- Alzheimer's Society 2013**
Ballard C, Burns A, Corbett A, Livingston G, Rasmussen J. *Helping you assess cognition: a practical toolkit for clinicians*. Alzheimer's Society, 2013.
- Alzheimer's Society 2014**
Prince M, Knapp M, Guerchet M, McCrone P, Prina M, Comas-Herrera M, et al. *Dementia UK: Update*. Alzheimer's Society (<http://www.alzheimers.org.uk/dementiauk>) 2014.
- American Psychiatric Association 1994**
American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-IV)*. Washington, DC: APA, 1994.
- American Psychiatric Association 2013**
American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington DC: American Psychiatric Publishing, 2013.
- Bahar-Fuchs 2013**
Bahar-Fuchs A, Clare L, Woods RT. Cognitive training and cognitive rehabilitation for mild to moderate Alzheimer's disease and vascular dementia. *Cochrane Database of Systematic Reviews* 2013, Issue 6. [DOI: 10.1002/14651858.CD003260.pub2]

- Beynon 2013**
Beynon R, Leeftang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database of Systematic Reviews* 2013, Issue 9. [DOI: 10.1002/14651858.MR000022.pub3]
- Birks 2006**
Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database of Systematic Reviews* 2006, Issue 1. [DOI: 10.1002/14651858.CD005593]
- Boustani 2003**
Boustani M, Peterson B, Hanson L, Harris R, Lohr KN. Screening for dementia in primary care: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 2003;**138**(11):927–37.
- Brayne 2012**
Brayne C, Davis D. Making Alzheimer's and dementia research fit for populations. *Lancet* 2012;**380**(9851):1441–3.
- Brunet 2012**
Brunet MD, McCartney M, Heath I, Tomlinson J, Gordon P, Cosgrove J, et al. There is no evidence base for proposed dementia screening. *BMJ (Clinical research ed.)* 2012; Vol. 345:e8588. [PUBMED: 23271709]
- Cochrane DTA Handbook Chapter 10**
Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y (editors). Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. Available from: <http://srdta.cochrane.org/> 2010.
- Cordell 2013**
Cordell CB, Borson S, Boustani M, Chodosh J, Reuben D, Verghese J, et al. Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimer's & Dementia* 2013;**9**(2):141–50. [PUBMED: 23265826]
- Davis 2013a**
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. [DOI: 10.1002/14651858.CD010460]
- Davis 2013b**
Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S. The Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementia disorders. *Cochrane Database of Systematic Reviews* 2013, Issue 10. [DOI: 10.1002/14651858.CD010775]
- Dementia CQUIN 2012**
Department of Health. Using the Commissioning for Quality and Innovation (CQUIN) payment framework: Guidance on new national goals for 2012-13. www.gov.uk/

government/uploads/system/uploads/attachment_data/file/215049/dh_133859.pdf.

Folstein 1975

Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;**12**(3):189–98.

Freitas 2012

Freitas S, Simoes M R, Maroco J, Alves L, Santana I. Construct Validity of the Montreal Cognitive Assessment (MoCA). *Journal of the International Neuropsychological Society* 2012;**18**(2):242–50.

Glanville 2012

Glanville JM, Cikaló M, Crawford F, Lowson P. Handsearching did not yield additional unique FDG-PET diagnostic test accuracy studies compared with electronic searches: a preliminary investigation. *Research Synthesis Methods* 2012 [Epub ahead of print].

Greaves 2010

Greaves I, Jolley D. National Dementia Strategy: well intentioned—but how well founded and how well directed? . *British Journal of General Practice* 2010;**60**(572):193–8. [PUBMED: 20202364]

Greening 2009

Greening L, Greaves I, Greaves N, Jolley D. Positive thinking on dementia in primary care: Gnosall Memory Clinic. *Community Practitioner* 2009;**82**(5):20–3. [PUBMED: 19480115]

Ismail 2010

Ismail Z, Rajji TK, Shulman KI. Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry* 2010;**25**:111–20.

Jack 2012

Jack CR Jr, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, et al. An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. *Annals of Neurology* 2012;**71**(6):765–75.

Leeflang 2008

Leeflang MM, Moons KG. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clinical Chemistry* 2008; Vol. 54, issue 4:729–37.

Lund 1994

The Lund and Manchester Groups. Clinical and neuropathological criteria for frontotemporal dementia. *Journal of Neurology, Neurosurgery, and Psychiatry* 1994;**57**(4):416–8.

Matthews 2002

Matthews FE, Denning T. Prevalence of dementia in institutional care. *Lancet* 2002;**360**(9328):225–6. [PUBMED: 12133659]

Matthews 2013

Matthews FE, Arthur A, Barnes LE, Bond J, Jagger C, Robinson L, Brayne C, Medical Research Council

Cognitive Function and Ageing Collaboration. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *Lancet* 2013;**382**(9902):1405–12. [DOI: 10.1016/S0140-6736(13)61570-6]

McKeith 2005

McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* 2005;**65**(12):1863–72.

McKhann 1984

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;**34**(7):939–44.

McShane 2006

McShane R, Areosa Sastre A, Minakaran N. Memantine for dementia. *Cochrane Database of Systematic Reviews* 2006, Issue 2. [DOI: 10.1002/14651858.CD003154.pub5]

Morris 1993

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;**43**(11):2412–4.

Nasreddine 2005

Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* 2005; **53**(4):695–9.

Noel-Storr 2014

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* 2014;**83**(4):364–73.

Plassman 2007

Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology* 2007;**29**(1-2):125–32. [PUBMED: 17975326]

RCPsych 2005

Royal College of Psychiatrists. *Who cares wins: improving the outcome for older people admitted to the general hospital*. Royal College of Psychiatrists, 2005.

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–90.

Román 1993

Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic

criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology* 1993;**43**(2):250–60.

Rutjes 2005

Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clinical Chemistry* 2005;**51**(8):1335–41. [MEDLINE: 15961549]

Sampson 2009

Sampson EL, Blanchard MR, Jones L, Tookman A, King M. Dementia in the acute hospital: prospective cohort study of prevalence and mortality. *British Journal of Psychiatry* 2009; **195**(1):61–6. [PUBMED: 19567898]

Whiting 2011a

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529–36.

Whiting 2011b

Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *Journal of Clinical Epidemiology* 2011;**64**(6):602–7.

Wimo 2010

Wimo A, Prince M. *World Alzheimer's Report 2010: the global economic impact of dementia*. London, UK: Alzheimer's Disease International (ADI), 2010.

Wittich 2010

Wittich W, Phillips N, Nasreddine ZS, Chertkow H. Sensitivity and specificity of the Montreal Cognitive Assessment modified for individuals who are visually impaired. *Journal of Visual Impairment and Blindness* 2010; **104**(6):360–8.

World Health Organization 1992

World Health Organisation. *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. Geneva: WHO, 1992.

Zekry 2009

Zekry D, Herrmann FR, Grandjean R, Vitale AM, De Pinho MF, Michel JB, et al. Does dementia predict adverse hospitalization outcomes? A prospective study in aged inpatients. *International Journal of Geriatric Psychiatry* 2009; **24**(3):283–91. [PUBMED: 18727147]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Cecato 2011

Study characteristics			
Patient sampling	Elderly participants assessed in geriatric clinic, not clear if consecutive or random		
Patient characteristics and setting	Geriatric clinic in Brazil. Prevalence of AD dementia = 28% 53 patients, 60 years and older, with at least 5 years of education (66% had 8 years or more) were submitted to MoCA Brazilian version. The diagnostic criteria for dementia were based on DSM-IV, NINCDS-ADRDA criteria were used for AD, and the Petersen criteria for MCI 15 participants were diagnosed with AD and 17 with MCI. Normal participants (n = 21) complained about memory problems but showed no evidence for dementia or MCI after neuropsychological assessment and neuroimaging		
Index tests	MoCA - Brazilian version.		
Target condition and reference standard(s)	AD Dementia, DSM IV.		
Flow and timing	No flow diagram available		
Comparative			
Notes	Extra information supplied by correspondence from author		
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?	Yes		
			Low
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			
Were sufficient data on MoCA application given for the test to be repeated in an independent study?	No			
				Unclear
DOMAIN 3: Reference Standard				
Is the reference standards likely to correctly classify the target condition?	Yes			
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes			
Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?	Unclear			
				Low
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?	Unclear			

Chen 2011

Study characteristics	
Patient sampling	Consecutive case series
Patient characteristics and setting	Location: Singapore National University Hospital, Singapore Prevalence = 54% N = 316 patients with mean age = 73 years, 87% Chinese, 53% female, with mean 6 years of education Memory clinic, consecutive patients over a 2 year period. Diagnosed using DSM-IV criteria
Index tests	MoCA-Singapore version
Target condition and reference standard(s)	AD Dementia, DSM IV, n = 172 (54%)
Flow and timing	None reported
Comparative	
Notes	Conference abstract with limited information, extra information supplied by correspondence from author

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?	Unclear		
			Low
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?	Unclear		

Chen 2011 (Continued)

Were sufficient data on MoCA application given for the test to be repeated in an independent study?	No			
				Unclear
DOMAIN 3: Reference Standard				
Is the reference standards likely to correctly classify the target condition?	Yes			
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear			
Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?	No			
				Low
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?	Unclear			

Kasai 2011

Study characteristics	
Patient sampling	Not clear from conference abstract, nor through contact with authors
Patient characteristics and setting	Community dwellers in Kurihara, Northern Japan Prevalence = 10%

Kasai 2011 (Continued)

	<p>N = 229 Dementia: n = 24, mean age = 82.6, mean 8.0 years of education MCI: n = 107, mean age = 80.8, mean 8.6 years of education Normal: n = 98 mean age = 78.8, mean 9.5 years of education Prevalence of dementia = 24/229 = 10.5%</p>		
Index tests	MoCA-Japanese		
Target condition and reference standard(s)	Dementia, Clinical Dementia Rating (CDR)		
Flow and timing	No flow diagram available		
Comparative			
Notes	Conference abstract with limited information, extra information supplied by correspondence from author		
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?	Unclear		
			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		
Were sufficient data on MoCA application given for the test to be repeated in an independent study?	No		

Kasai 2011 (Continued)

				Unclear
DOMAIN 3: Reference Standard				
Is the reference standards likely to correctly classify the target condition?	Yes			
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes			
Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?	Yes			
				Low
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?	Unclear			

Larner 2012

Study characteristics

Patient sampling	Consecutive case series
Patient characteristics and setting	Memory clinic in Liverpool, UK Prevalence dementia = 24% N = 150 Sample characteristics: 62% male, median age = 61 years (range = 20 to 87 years) Educational level not reported Consecutive new patient referrals prospectively recruited from a memory clinic in regional neuroscience centre

Index tests	MoCA		
Target condition and reference standard(s)	AD Dementia, DSM IV. n = 36 (24%) MCI, Petersen criteria. n = 29 (19%) Non-cases: n = 85 (57%)		
Flow and timing	None		
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?	Unclear		
			Low
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		
Were sufficient data on MoCA application given for the test to be repeated in an independent study?	No		
			Low
DOMAIN 3: Reference Standard			

Larner 2012 (Continued)

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?	Yes		
			Low
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?	Unclear		

Lee 2008

Study characteristics	
Patient sampling	All participants were aged 65 and over and were recruited from among attenders at a community dementia care centre in Seoul where people without dementia can request memory testing. Most of the non-dementia participants were recruited from this centre. AD patients were registered outpatients at Seoul National Hospital but it is not clear whether they had also attended the community dementia care centre. Thus sampling frame is unclear
Patient characteristics and setting	Location: Community dementia care centre and Seoul National University Boramae Hospital, Korea Prevalence AD dementia = 22% N = 196, 44 with AD, 37 with MCI, and 115 normal subjects (N) Mean age: 70.4 in AD, 71.3 in MCI, 69.1 in N % female: 52% in AD, 62% in MCI, 70% in N Mean years education: 7.9 in AD, 8.3 in MCI, 8.0 in N

Index tests	MoCA-Korean		
Target condition and reference standard(s)	Dementia, CDR		
Flow and timing	No flow diagram available		
Comparative			
Notes	Extra information supplied by author		
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		
			Unclear
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		
Were sufficient data on MoCA application given for the test to be repeated in an independent study?	Yes		
			Low
DOMAIN 3: Reference Standard			

Lee 2008 (Continued)

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?	Unclear		
			Low
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?	Unclear		

Lu 2011

Study characteristics	
Patient sampling	Multistage stratified cluster sampling of community-dwelling adults aged 65 years and older. Study based on 2 samples from urban and rural areas
Patient characteristics and setting	China N = 8411; mean age = 73.0 years (range = 65 to 103), 53.7% female, Mean years of education = 5.8 (range = 0 to 22) Cognitively normal: n = 6283, mean age = 72.0 (range = 65 to 100), 52.1% female, Mean years of education = 6.7 (range = 0 to 22) Dementia; n = 441, mean age = 78.9 years (range = 65 to 103), female = 68.7%, Mean years of education = 2.5 (range = 0 to 19). Prevalence of dementia = 441/8411 = 5.2% Inclusion criteria were (1) 65 years old and older, (2) Han Chinese, (3) listed in the census of the community registry office and living in the target community for at least 1 year preceding the survey date, and (4) free of comorbid conditions that could affect assessment

	Those listed in the census but institutionalised were not included in the study		
Index tests	MoCA-Beijing version		
Target condition and reference standard(s)	Dementia, DSM-IV AD dementia, NINCDS-ADRDA		
Flow and timing			
Comparative			
Notes	Results presented were adjusted for education		
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?	Unclear		
			Low
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		
Were sufficient data on MoCA application given for the test to be repeated in an independent study?	Yes		
			Low
DOMAIN 3: Reference Standard			

Lu 2011 (Continued)

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?	Yes		
			Low
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?	Unclear		

Smith 2007

Study characteristics	
Patient sampling	case series
Patient characteristics and setting	<p>Location: UK N = 67 Study sample characteristics: Mean age = 73.6 years, 49.3% female, mean level of education = 12.1years Memory assessment research centre All patients scoring > 24 on MMSE Non-cases were patients referred to memory clinic and subsequently diagnosed with psychiatric illness that explained the subjective memory complaints or who demonstrated no objective evidence of memory loss (CAMCOG score > 89 points)</p>
Index tests	MoCA

Smith 2007 (Continued)

Target condition and reference standard(s)	Dementia, ICD-10 criteria (AD = 18, vascular dementia = 13, Parkinson's disease dementia = 1). n = 32 (48%) MCI, Petersen's criteria. n = 23 Non-cases: n = 12		
Flow and timing	6 months		
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?	Unclear		
			Unclear
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		
Were sufficient data on MoCA application given for the test to be repeated in an independent study?	Unclear		
			Unclear
DOMAIN 3: Reference Standard			

Smith 2007 (Continued)

Is the reference standards likely to correctly classify the target condition?	Yes			
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes			
Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?	Unclear			
				Low
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?	Unclear			
Were all patients included in the analysis?	Unclear			

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Aulwes 2012	unable to calculate specificity
Chang 2012	MCI population only
Costa 2012	case control study
Fujiwara 2010	case control study
Guo 2010	case control study

(Continued)

Kandiah 2011	unable to calculate sensitivity or specificity
Karunaratne 2011	case control study
Luis 2009	sampled from varied populations
Magierska 2012	case control study
Nasreddine 2005	case control study
Selekler 2009	case control study
Thissen 2010	case control study
Tsai 2012	case control study
Zhang 2008	sampled from nursing home population

DATA

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

Tests. Data tables by test

Test	No. of studies	No. of participants
1 MoCA Secondary care	5	745
2 MoCA community based studies	2	8640

Test 1. MoCA Secondary care.

Review: Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias

Test: 1 MoCA Secondary care

Study	TP	FP	FN	TN	Sensitivity	Specificity	Sensitivity	Specificity
Cecato 2011	14	5	1	33	0.93 [0.68, 1.00]	0.87 [0.72, 0.96]		
Chen 2011	162	49	10	95	0.94 [0.90, 0.97]	0.66 [0.58, 0.74]		
Lamer 2012	63	34	2	51	0.97 [0.89, 1.00]	0.60 [0.49, 0.70]		
Lee 2008	44	56	0	59	1.00 [0.92, 1.00]	0.51 [0.42, 0.61]		
Smith 2007	30	17	2	18	0.94 [0.79, 0.99]	0.51 [0.34, 0.69]		

Test 2. MoCA community based studies.

Review: Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias

Test: 2 MoCA community based studies

Study	TP	FP	FN	TN	Sensitivity	Specificity	Sensitivity	Specificity
Kasai 2011	101	42	30	56	0.77 [0.69, 0.84]	0.57 [0.47, 0.67]		
Lu 2011	2094	3003	34	3280	0.98 [0.98, 0.99]	0.52 [0.51, 0.53]		

APPENDICES

Appendix I. Sources searched and search strategies

Search narrative: for each of the major healthcare databases a single concept search using only the index test was performed. This was felt to be the simplest and most sensitive approach.

Source	Search strategy	Hits retrieved
1. MEDLINE In-Process and other non-indexed citations and MEDLINE 1950 to August 2012 (Ovid SP)	1. "montreal cognitive assessment*" .mp. 2. MoCA.mp. 3. 1 or 2	265
2. EMBASE 1974 to 2012 August 31 (Ovid SP)	1. MoCA.mp. 2. "montreal cognitive assessment*" .mp. 3. 1 or 2	625
3. PsycINFO 1806 to August week 4 2012 (Ovid SP)	1. MoCA.mp. 2. "montreal cognitive assessment*" .mp. 3. 1 or 2	218
4. Thomson Reuters Web of Science: Biosis previews 1926 to August 2012 (ISI Web of Knowledge)	Topic = (MoCA OR "montreal cognitive assessment*") Timespan = All Years. Databases = BIOSIS Previews. Lemmatization = On	274

(Continued)

5. Thomson Reuters Web of Science: Web of Science Core Collection (including Conference Proceedings Citation Index) (1945 to August 2012)	Topic = (MoCA OR “montreal cognitive assessment*”) Timespan = All Years. Databases = SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Lemmatization = On	538
6. LILACS (BIREME) to August 2012	MoCA OR montreal cognitive assessment [Words]	33
7. ALOIS (CDCIG register) (see below for detailed explanation of what is contained within the ALOIS register)	MoCA OR “montreal cognitive assessment”	22
TOTAL before de-duplication		1942
TOTAL after de-duplication		1122

In addition to the above single concept search based on the Index test, the Cochrane Dementia and Cognitive Improvement Group run a more complex, multi-concept search in several healthcare databases each month primarily for the identification of diagnostic test accuracy studies of neuropsychological tests. Where possible the full texts of the studies identified are obtained. This approach is expected to help identify those papers where the index test of interest (in this case MoCA) is used and the paper contains usable data but where MoCA was not specifically alluded to in the parts of the electronic record available for search retrieval.

The strategy used is below:

The MEDLINE search uses the following concepts:

A Specific neuropsychological tests

B General terms (both free text and MeSH) for tests/testing/screening

C Outcome: dementia diagnosis (unfocused MeSH with diagnostic sub-headings)

D Condition of interest: Dementia (general dementia terms both free text and MeSH - exploded and unfocused)

E Methodological filter: NOT used to limit all search

Concept combination:

1. (A OR B) AND C

2. (A OR B) AND D AND E

3. A AND E

= 1 OR 2 OR 3

Setting is not included as a concept in the MEDLINE search as these terms are generally not indexed well or consistently. This means that the search has been kept deliberately sensitive by not restricting it to a particular setting.

The search strategy

1. “word recall”.ti,ab.

2. (“7-minute screen” OR “seven-minute screen”).ti,ab.

3. (“6 item cognitive impairment test” OR “six-item cognitive impairment test”).ti,ab.

4. “6 CIT”.ti,ab.

5. “AB cognitive screen”.ti,ab.

6. “abbreviated mental test”.ti,ab.

7. “ADAS-cog”.ti,ab.

8. AD8.ti,ab.

9. “inform* interview”.ti,ab.

10. “animal fluency test”.ti,ab.

11. "brief alzheimer* screen".ti,ab.
12. "brief cognitive scale".ti,ab.
13. "clinical dementia rating scale".ti,ab.
14. "clinical dementia test".ti,ab.
15. "community screening interview for dementia".ti,ab.
16. "cognitive abilities screening instrument".ti,ab.
17. "cognitive assessment screening test".ti,ab.
18. "cognitive capacity screening examination".ti,ab.
19. "clock drawing test".ti,ab.
20. "deterioration cognitive observee".ti,ab.
21. ("Dem Tect" OR DemTect).ti,ab.
22. "object memory evaluation".ti,ab.
23. "IQCODE".ti,ab.
24. "mattis dementia rating scale".ti,ab.
25. "memory impairment screen".ti,ab.
26. "minnesota cognitive acuity screen".ti,ab.
27. "mini-cog".ti,ab.
28. "mini-mental state exam*".ti,ab.
29. "mmse".ti,ab.
30. "modified mini-mental state exam".ti,ab.
31. "3MS".ti,ab.
32. "neurobehavio?ral cognitive status exam*".ti,ab.
33. "cognistar".ti,ab.
34. "quick cognitive screening test".ti,ab.
35. "QCST".ti,ab.
36. "rapid dementia screening test".ti,ab.
37. "RDST".ti,ab.
38. "repeatable battery for the assessment of neuropsychological status".ti,ab.
39. "RBANS".ti,ab.
40. "rowland universal dementia assessment scale".ti,ab.
41. "rudas".ti,ab.
42. "self-administered gerocognitive exam*".ti,ab.
43. ("self-administered" and "SAGE").ti,ab.
44. "self-administered computerized screening test for dementia".ti,ab.
45. "short and sweet screening instrument".ti,ab.
46. "sassi".ti,ab.
47. "short cognitive performance test".ti,ab.
48. "syndrome kurztest".ti,ab.
49. ("six item screener" OR "6-item screener").ti,ab.
50. "short memory questionnaire".ti,ab.
51. ("short memory questionnaire" and "SMQ").ti,ab.
52. "short orientation memory concentration test".ti,ab.
53. "s-omc".ti,ab.
54. "short blessed test".ti,ab.
55. "short portable mental status questionnaire".ti,ab.
56. "spmsq".ti,ab.
57. "short test of mental status".ti,ab.
58. "telephone interview of cognitive status modified".ti,ab.
59. "tics-m".ti,ab.
60. "trail making test".ti,ab.
61. "verbal fluency categories".ti,ab.
62. "WORLD test".ti,ab.
63. "general practitioner assessment of cognition".ti,ab.

64. "GPCOG".ti,ab.
65. "Hopkins verbal learning test".ti,ab.
66. "HVLt".ti,ab.
67. "time and change test".ti,ab.
68. "modified world test".ti,ab.
69. "symptoms of dementia screener".ti,ab.
70. "dementia questionnaire".ti,ab.
71. "7MS".ti,ab.
72. ("concord informant dementia scale" or CIDS).ti,ab.
73. (SAPH or "dementia screening and perceived harm").ti,ab.
74. or/1-73
75. exp Dementia/
76. Delirium, Dementia, Amnesic, Cognitive Disorders/
77. dement*.ti,ab.
78. alzheimer*.ti,ab.
79. AD.ti,ab.
80. ("lewy bod*" or DLB or LBD or FTD or FTLD or "frontotemporal lobar degeneration" or "frontaltemporal dement*).ti,ab.
81. "cognit* impair".ti,ab.
82. (cognit* adj4 (disorder* or declin* or fail* or function* or degenerat* or deteriorat*)).ti,ab.
83. (memory adj3 (complain* or declin* or function* or disorder*)).ti,ab.
84. or/75-83
85. exp "sensitivity and specificity"/
86. "reproducibility of results"/
87. (predict* adj3 (dement* or AD or alzheimer*)).ti,ab.
88. (identif* adj3 (dement* or AD or alzheimer*)).ti,ab.
89. (discriminat* adj3 (dement* or AD or alzheimer*)).ti,ab.
90. (distinguish* adj3 (dement* or AD or alzheimer*)).ti,ab.
91. (differenti* adj3 (dement* or AD or alzheimer*)).ti,ab.
92. diagnos*.ti.
93. di.fs.
94. sensitivit*.ab.
95. specificit*.ab.
96. (ROC or "receiver operat*").ab.
97. Area under curve/
98. ("Area under curve" or AUC).ab.
99. (detect* adj3 (dement* or AD or alzheimer*)).ti,ab.
100. sROC.ab.
101. accura*.ti,ab.
102. (likelihood adj3 (ratio* or function*)).ab.
103. (conver* adj3 (dement* or AD or alzheimer*)).ti,ab.
104. ((true or false) adj3 (positive* or negative*)).ab.
105. ((positive* or negative* or false or true) adj3 rate*).ti,ab.
106. or/85-105
107. exp dementia/di
108. Cognition Disorders/di [Diagnosis]
109. Memory Disorders/di
110. or/107-109
111. *Neuropsychological Tests/
112. *Questionnaires/
113. Geriatric Assessment/mt
114. *Geriatric Assessment/
115. Neuropsychological Tests/mt, st
116. "neuropsychological test*".ti,ab.

117. (neuropsychological adj (assess* or evaluat* or test*)).ti,ab.
118. (neuropsychological adj (assess* or evaluat* or test* or exam* or battery)).ti,ab.
119. Self report/
120. self-assessment/ or diagnostic self evaluation/
121. Mass Screening/
122. early diagnosis/
123. or/111-122
124. 74 or 123
125. 110 and 124
126. 74 or 123
127. 84 and 106 and 126
128. 74 and 106
129. 125 or 127 or 128
130. exp Animals/ not Humans.sh.
131. 129 not 130

Appendix 2. Information for extraction to pro forma

Bibliographic details of primary paper:

- Author, title of study, year and journal

Details of index test:

- Method of [index test] administration, including who administered and interpreted the test, and their training
- Thresholds used to define positive and negative tests

Reference Standard:

- Reference standard used
- Method of [reference standard] administration, including who administered the test and their training

Study population:

- Number of participants
- Age
- Gender
- Other characteristics e.g. ApoE status
- Settings: i) community; ii) primary care; iii) secondary care outpatients; iv) secondary care inpatients and residential care
- Participant recruitment
- Sampling procedures
- Time between index test and reference standard
- Proportion of people with dementia in sample
- Subtype and stage of dementia if available
- MCI definition used (if applicable)
- Duration of follow-up in delayed verification studies
- Attrition and missing data

Appendix 3. Assessment of methodological quality QUADAS-2

DOMAIN	PARTICIPANT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING
Description	Describe methods of participant selection: Describe included participants (prior testing, presentation, intended use of index test and setting) :	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any participants who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram): Describe the time interval and any interventions between index test(s) and reference standard:
Signalling questions (yes/no/unclear)	Was a consecutive or random sample of participants enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided?	If a threshold was used, was it prespecified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all participants receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all participants receive the same reference standard?
				Were all participants included in the analysis?
Risk of bias: (High/low/ unclear)	Could the selection of participants have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the participant flow have introduced bias?
Concerns regarding applicability: (High/low/ unclear)	Are there concerns that the included participants do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Appendix 4. Anchoring statements for quality assessment of MoCA diagnostic studies

We provide some core anchoring statements for quality assessment of diagnostic test accuracy reviews of the MoCA 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' then the risk of bias can be judged to be 'low'. If a question is answered 'no' this indicates a potential risk of 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 less important for assessing overall quality. To assist, we describe a 'weighting' system. Where an item is weighted 'high risk' then 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 index test is fundamental. If this blinding was not present then the item on reference standard should be scored 'high risk of bias', regardless of the other contributory elements. Where an item is weighted 'low risk' then 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 will be judged on whether other signalling questions (with a high risk of bias) for the same domain are also answered 'no'.

In assessing individual items, the score of unclear should only be given if there is genuine uncertainty. In these situations review authors will contact the relevant study teams for additional information.

Anchoring statements to assist with assessment for risk of bias

Domain 1: Participant selection

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

Was a consecutive or random sample of participants enrolled?

Where sampling is used, the methods least likely to cause bias are consecutive sampling or random sampling, which should be stated and/or described. 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 sometimes they are the only studies available, especially if the index test is expensive and/or invasive. Nested case-control designs (systematically selected from a defined population cohort) are less prone to bias but they will still narrow the spectrum of participants that receive the index test. Study designs (both cohort and case-control) that may also increase bias are those designs where the study team deliberately increase or decrease 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). Where exclusions are detailed, the study will be graded as 'low risk' if exclusions are felt to be appropriate by the review authors. Certain exclusions common to many studies of dementia are: medical instability; terminal disease; alcohol/substance misuse; concomitant psychiatric diagnosis; 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. Post hoc exclusions will be labelled '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, setting will be particularly important - the review authors should consider population in terms of symptoms; pre-testing; potential disease prevalence. Studies that use very selected subjects or subgroups will be classified 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, and full details of the blinding procedure are not required. This item may be scored as 'low risk' if explicitly described or if there is a clear temporal pattern to the order of testing that precludes the need for formal blinding, i.e. 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 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 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 and 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 (for example self-completed questionnaire versus direct questioning interview); nature of informant; 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 there should be evidence of validation.

Weighting: Low risk

Applicability: are there concerns that the index test, its conduct, or interpretation 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 vary 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 to assist with clinical diagnosis of dementia include those detailed in DSM-IV and ICD-10. Criteria specific to dementia subtypes include but are not limited to NINCDS-ADRDA criteria for Alzheimer's dementia; McKeith criteria for Lewy Body dementia; Lund criteria for frontotemporal dementias; and the NINDS-AIREN criteria for vascular dementia. Where the criteria used for assessment are not familiar 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, and full details of the blinding procedure are not required. This may be scored as 'low risk' if explicitly described or if there is a clear temporal pattern to order of testing, i.e. 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 is a usual component 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/expertise of the assessor; and whether additional information was available to inform the diagnosis (e.g. neuroimaging; other neuropsychological test results), 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 is the possibility that some methods of dementia assessment, although valid, may diagnose a far smaller or larger proportion of participants with disease than in usual clinical practice. In this instance 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 reference standard?

For a cross sectional study design, there is potential for the subject to change between assessments, however dementia is a slowly progressive disease, which is not reversible. The ideal scenario would be a same day assessment, but longer periods of time (for example, 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 subjects receive the same reference standard?

There may be scenarios where those who score 'test positive' on the index test have a more detailed assessment for the target condition. Where dementia assessment (or 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 due to mortality, and it is likely to be greater in participants with the target condition. Drop-outs (and missing data) should be accounted for. Attrition that is higher than expected (compared to other similar studies) should be treated as at high risk of bias. We have defined a cut-off of greater than 20% attrition as being high risk but this will be highly dependent on the length of follow-up in individual studies.

Weighting: High risk

CONTRIBUTIONS OF AUTHORS

DD: designed and drafted protocol; study selection and data extraction, Characteristics of included and excluded studies tables; entered data and data entry check, QUADAS-2 assessment; updated methods, drafted results, findings, discussion and authors' conclusions; finalised the review manuscript.

STC: designed and drafted protocol; study selection, updated methods, drafted results, findings, discussion and authors' conclusions; finalised the review manuscript.

JY: Characteristics of included and excluded studies tables; entered data and data entry check; set up 'Data and analyses' tables; drafting of manuscript.

ANS: writing the search strategy, searching and initial screening of search results.

CB: supervisory overview of protocol, finalising review manuscript.

SC: designed and drafted protocol; study selection and data extraction, Characteristics of included and excluded studies tables; entered data and data entry check, QUADAS-2 assessment; updated methods, drafted results, findings, discussion and authors' conclusions; finalised the review manuscript.

DECLARATIONS OF INTEREST

None.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- NIHR, UK.

This review was supported by the National Institute for Health Research, via a Cochrane Programme Grant to the Cochrane Dementia and Cognitive Improvement Group. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Our protocol specified that we would include cross-sectional cohort studies (i.e. studies that examined unselected samples), but to provide for the possibility of not retrieving any studies of these designs, we stated we would also consider including case-control studies (which are considered to be at high risk of bias); the latter was not necessary.

We had intended to carry out a meta-analysis of included studies but, due to differences in study design, participant sampling, reference standards and reported thresholds, we judged that a meta-analysis was not appropriate. We conducted a narrative review instead.