Article:
Ozer, S, Young, J, Champ, C et al. (1 more author) (2016) A systematic review of the diagnostic test accuracy of brief cognitive tests to detect amnestic mild cognitive impairment. International Journal of Geriatric Psychiatry, 31 (11). pp. 1139-1150. ISSN 0885-6230

https://doi.org/10.1002/gps.4444

Title: A systematic review of the diagnostic test accuracy of brief cognitive tests to detect amnestic mild cognitive impairment

Authors:

Seline Ozer[1, 2], John Young[1], Claire Champ[2], Melanie Burke[2]

[1] Academic Unit of Elderly Care and Rehabilitation, University of Leeds, Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Duckworth Lane, Bradford, West Yorkshire, UK
[2] School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, UK

Key words:

Ageing, Alzheimer's disease, Dementia, MCI, SRR

Key Points:

- Recognition of Mild Cognitive Impairment (MCI) is becoming increasingly important.
- A large number of brief cognitive tests for identifying the amnestic form of MCI (aMCI) have been developed.
- The Montreal Cognitive Assessment (MoCA) is the most comprehensively investigated, with high sensitivity and good test-retest reliability.
- Lack of evidence on predictive validity, and concerns over quality of studies limit confidence. Future high-quality studies that use unbiased sampling methods are required to further validate some of the most promising brief cognitive tests.

Word count: 4000

Abstract

Objective: People with amnestic Mild Cognitive Impairment (aMCI) are at an increased risk of developing dementia. Efficient ways of identifying this "at risk" population are required for larger scale research studies. This systematic review describes the diagnostic accuracy of brief cognitive tests for detecting aMCI.

Methods: Fifteen databases were searched from 1999 to July 2013 to identify papers for inclusion. Prospective studies assessing the diagnostic test accuracy of simple and brief cognitive tests for identifying people with aMCI against a reference standard (Petersen criteria) were included. Sensitivity, specificity, positive and negative predictive values and likelihood ratios were calculated. Predictive validity and test-retest reliability were also extracted, when provided. Risk of bias was assessed using the QUADAS-2 tool.

Results: Thirty nine studies assessing 42 index tests were included. The MoCA was the most comprehensively assessed test with evidence of high sensitivity for aMCI and good test-retest reliability, but low specificity reported by the only study judged to be at low risk of bias. Other brief cognitive tests that include an assessment of word recall, and multi-task tests that assess several cognitive domains, were also found to exhibit high sensitivities and reasonable specificities. However, the confidence of the findings was affected by overall low quality of the contributing studies.

Conclusion: Several brief cognitive tests have shown promising diagnostic test accuracy results for identifying aMCI. However, concerns over the quality of the constituent studies and lack of evidence on the predictive validity of these tests, mean that new validation studies are warranted.

Introduction

Mild Cognitive Impairment (MCI) is a term used to describe the transitional state between normal aging and established dementia (Petersen et al., 1999). The international definition of MCI is "cognitive decline greater than that expected for an individual's age and education level but that does not interfere notably with activities of daily life" (Gauthier et al., 2006).

The most widely used current procedure for diagnosing MCI is based on the Petersen criteria (Petersen et al., 1999). The original criteria focussed on memory impairment, but during recent years they have been expanded to define subtypes including non-amnestic (without memory impairment), as well as single and multi-domain impaired forms (Petersen, 2004). MCI of the amnestic-type (both single and multi-domain) is the focus of this review since it is the commonest subtype. A 2.6 fold increased incidence rate of aMCI compared to non-amnestic MCI (naMCI) was reported in a large longitudinal study (Roberts et al., 2012), and aMCI associated with elevated rates of conversion to Alzheimer's disease (Petersen et al., 2001), the most common form of dementia (Ferri et al., 2005).

It has been suggested that targeting interventions on people with aMCI might prevent or slow the decline into dementia (Petersen et al., 2009). However, the application of the Petersen criteria to diagnose aMCI is not straightforward and requires assessment by a trained specialist, along with considerable commitment from the patient to complete a complex battery of cognitive tests that are time consuming and can be fatiguing. Less demanding cognitive tests might have utility to provide a more efficient method to identify people with aMCI in research or clinical settings, but this would require test accuracy to have been confirmed.

A systematic review was therefore conducted to identify the brief cognitive tests that have been used to identify people with aMCI and to evaluate the evidence for their accuracy. The work updates an earlier review (Lonie et al., 2009) but is also more focussed, including only those tools that take less than 15 minutes to administer and incorporates a quality appraisal of the included studies, an aspect not covered in the previous review. In addition, new information on any predictive validity and reliability measures reported for the tools is included.

Methods

A systematic review was performed to describe the test accuracy of brief cognitive tests that have been used to identify people with aMCI. The methodology and reporting of this review followed standard guidance (Deeks et al., 2010, Moher et al., 2009).

Criteria for considering studies for this review

Prospective studies assessing the DTA of brief and simple cognitive tests used to identify people with aMCI (index tests) against a reference standard were considered for inclusion. Studies published in a language other than English were excluded. Only peer reviewed articles were included.

Participants

Participants were people with aMCI (single and multi-domain) diagnosed according to the Petersen criteria.

Index tests

Index tests considered for inclusion were those that were considered to be brief and simple cognitive tests, where "brief" was defined as: (1) taking less than 15 minutes to administer and "simple" was defined as: (2) not computer-based or requiring specialist equipment; and (3) not requiring specialist staff for administration. Studies assessing telephone administered or wholly carer/informant rated screening tools were excluded.

Reference standard

As it is the most widely used procedure for diagnosing aMCI, only those studies which used the Petersen criteria as the reference standard for verification of diagnosis were included in the review.

Search methods for identification of studies

The following databases were searched to identify studies for inclusion: MEDLINE, EMBASE, BIOSIS Previews, Web of Science, PsychINFO, LILACS, CINAHL, AMED, Cochrane Library, ASSIA, IBSS, PsychARTICLES, Scopus, Sociological Abstracts and ProQuest dissertations and theses. Databases were searched from 1999 – July 2013 (see Appendix 2 for the strategy used to search CINAHL).

Selection of studies

Two independent reviewers assessed all titles and full-text articles for inclusion. Any disagreements were resolved by consensus.

Data extraction and management

Two independent reviewers extracted all data using a standardised data extraction form and any disagreements were resolved by consensus. The extracted data included information on

methods (e.g. study design, recruitment procedure) and participant characteristics. The operationalised reference criteria used and index tests, including cut-off points for diagnosis, were recorded and any validity and test-retest reliability measures reported for the index tests were also extracted. The measures for DTA included: sensitivity, specificity and AUC (area under the receiver operating characteristic (ROC) curve) for discriminating between aMCI and cognitively normal participants. Where sensitivity/specificity was reported for more than one cut-off, only that reported as optimal by the author was extracted. In cases where no optimal value was stated, the cut-off providing the highest sensitivity was extracted. Where available, data concerning the ability of the test to predict future dementia were also extracted.

Assessment of methodological quality

Two independent reviewers assessed the methodological quality of the included studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool (Whiting et al., 2011), as recommended by The Cochrane Collaboration. QUADAS-2 involves a structured assessment using signalling questions in four domains: patient selection, index test, reference standard, and flow and timing.

Statistical analysis and data synthesis

RevMan 5.2 software (http://tech.cochrane.org/revman) was used to construct 2 x 2 tables of index test performance (i.e. number of true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN)) using reported sensitivity and specificity values, as well as total number of participants and proportion of aMCI participants. These data were used to calculate the sensitivity and specificity with 95% confidence intervals (CI) and construct forest plots. Positive and negative predictive values (PPVs and NPVs) and positive

and negative likelihood ratios (LR+ and LR-) were also calculated. Meta-analysis was performed, where appropriate, using STATA version 13.0 (StatCorp, 2013) software. Where sufficient studies were found ($n \geq 4$) that reported DTA data for the same test and cut-off, pooled estimates of sensitivity, specificity, LR+, LR- and summary diagnostic odds ratios were produced using a random effects bivariate model (Harbord et al., 2007, Reitsma et al., 2005) and heterogeneity was assessed using the I^2 statistic. Where four or more studies were assessing the same test at different cut-offs, summary ROC curves were produced using the STATA MIDAS module (Dwamena et al., 2010).

Results

Results of the search

The search identified 6431 citations, of which, 158 were considered as potentially relevant and the full articles obtained. Subsequently, 119 reports were excluded and 39 included (see Figure 1). Of these, 37 were cross-sectional DTA studies and two were longitudinal DTA studies that investigated the predictive validity of several index tests. There were 5766 aMCI and cognitively normal participants included in these studies. The mean prevalence of aMCI reported in the cross-sectional DTA studies was 42.4% but varied considerably from 3.1% to 72%. The majority of these studies (n = 32) recruited their aMCI population from secondary care settings, such as memory clinics and hospital departments, four studies recruited from the community and the remainder recruited from a mixture of secondary care and community based settings (n=3). The characteristics of the included studies are summarised in Table 1.

[insert Figure 1 here]

[insert Table 1 here]

In total, 42 brief and simple cognitive tests were investigated in the 39 studies. These tests are listed and described in Supplementary data, Appendix 3. Thirty five of the index tests involve single tasks. Of these, eighteen involve a memory component where either words, shapes or digits need to be remembered and recalled. Three of the memory tasks (one of which has four scoring methods) are verbal learning tasks (VLTs) and have been grouped together. These tasks involve the participant recalling a word list after it has been read out to them. The remaining seventeen tasks involve the testing of other, non-memory cognitive domains, such as semantic knowledge, visuospatial processing, attention/orientation and executive function/fluency. Within this grouping, there are four versions of clock drawing tests (CDTs) and five verbal fluency tasks (VFTs) The CDTs involve the participant drawing a clock and setting the time. The VFTs involve the participant naming as many words as they can within a certain time period (usually one minute), either from a certain category or beginning with a certain letter. Seven of the index tests involve multiple tasks. By their nature, they tend to involve the assessment of more cognitive domains (all involve a memory component) and take slightly longer to administer.

Findings

DTA assessment

A summary of the DTA results reported for identifying people with aMCI is presented in Table 2 for single task index tests (grouped as memory and non-memory tasks) and Table 3 for multi-task index tests. Forest plots of sensitivity and specificity for each index test are presented in Supplementary data, Appendix 4 and 5. Across the studies, sensitivity ranged widely from 7 to 100%, and specificity from 35 to 100%.

[insert Table 2 here]

[insert Table 3 here]

Of the single task index tests involving a memory component, the Auditory Verbal Learning Test- Short Delay Recall task (AVLT-SR) showed the highest sensitivity (97%) for detecting aMCI, with a high specificity also (73%). The high AUC value (0.94) confirms the high diagnostic accuracy of the test. The Consortium to Establish a Registry for Alzheimer's Disease- Word List Memory Test Recognition Discrimination Index task (CERAD-WLREDI) also showed high sensitivity at 94% but very low specificity (35%). The delayed recall task from the same Word List Memory Test (CERAD-WLDR) showed high sensitivity in two of the studies (82%-83%) but low in one study (27%) and the Hopkins Verbal Learning Test- Wordlist Learning task (HVLT-LE) showed high sensitivity across two studies (83-84%). The Florida Brief Memory Screen task (FBMS), which also involves word recall, also showed high sensitivity (83%) and specificity (88%).

The non-memory single task index tests in general demonstrated lower sensitivity for detecting aMCI than those involving a memory component. The evidence of accuracy for the VFTs was uncertain, with sensitivities ranging from 27% to 83%. There were five studies that investigated four different versions of the CDT . Again, results were fairly inconsistent across the studies with sensitivities ranging between 7% and 76%.

Of the multi-task index tests, the Mini Mental State Examination Scale (MMSE) was the most frequently investigated (17 studies). Sensitivity was reported for a number of cut-off values. The highest sensitivity reported was 76% for a cut-off value of ≤26. However, sensitivities were generally lower for other cut-offs, with most studies reporting sensitivities between 13% and 68%. The next most frequently investigated multi-task index test was the

Montreal Cognitive Assessment (MoCA) with results reported in 13 studies. Again, sensitivity was reported for a number cut-offs and the highest sensitivity reported was 100% at a cut-off of <24 but specificity was low at 50%. Sensitivities ranged from 81% to 96% for other reported cut-off values. Four studies reported sensitivity and specificity for the same cut-off of 23/24 and thus were combined in a meta-analysis (see Supplementary data, Appendix 6). A pooled sensitivity and specificity of 86% was calculated at this cut-off value across the four studies (see Table 3).

Summary ROC curves were produced to provide a visual summary of the DTA reported across all studies investigating the VFT-Animals, MMSE and MoCA (see Figure 2). These curves illustrate that MoCA generally performed with higher sensitivity/specificity across studies than the VFT-Animals and MMSE, with most points gathering towards the top left hand corner of the ROC space. The high AUC value for the MoCA sROC curve (0.92, 95%CI 0.89-0.94) also confirms the higher diagnostic accuracy of the test in comparison with the VFT-Animals and MMSE, which had AUC values of 0.75 (95%CI 0.71-0.79) and 0.73 (95%CI 0.69-0.77) respectively.

[insert Figure 2 here]

Sensitivity was fairly high for all other multi-task index tests (ranging from 76% to 96%). Of these, the Memory Alteration Test (M@T) showed the highest sensitivity (96%) and fairly high specificity (70 – 79%). The high reported AUC values (0.88 - 0.93) confirm the high diagnostic accuracy of this test.

Four studies reported DTA for combinations of index tests (see Supplementary data, Appendix 7 and 8). The combination of MMSE and CDT-command showed the highest sensitivity for aMCI (76%).

Predictive Validity

Two longitudinal studies reported on the validity of 12 index tests for predicting future dementia over periods of one (Ahmed et al., 2008) or three (Sarazin et al., 2007) years (see Supplementary data, Appendix 9 and 10). The Free and Cued Selective Recall Reminding Test- Total Recall task (FCSRT-Total Recall) was the most accurate prognostic test with a sensitivity of 80% for identifying progressors and a specificity of 90% for identifying non-progressors. All other tests showed relatively low sensitivities.

Test-Retest Reliability

Test-retest reliability was available for seven of the index tests (see Supplementary data, Appendix 11). The MoCA was the most frequently investigated (eight studies). Most studies assessed reliability using the Intraclass Correlation Coefficient (ICC) with values ranging from 0.75-0.92 indicating a fairly high to a high reliability over a range of time periods from four weeks to 18 months. Reported ICCs for the MMSE tended to be slightly lower ranging from 0.67-0.76. High test-retest reliability was reported in one study for the A Quick Test of Cognitive Speed- colour-form naming task (AQT-CF) (ICC = 0.88). For other index tests, namely CDT, Florida Brief Memory Screen (FBMS) and MMSE & CDT, reliability data reporting was incomplete without a description of the method used to assess reliability.

Methodological quality

Of the 39 studies, 27 were assessed as high risk of bias; 11 were unclear risk of bias; and only one study (McLennan et al., 2011) scored a low risk of bias across all domains assessed (see Table 1). A summary of the quality assessment results across all four QUADAS-2 domains is provided in Figure 3. Most studies (n=22) were judged to be at a high risk of bias in the patient selection domain due to unblinding of the participant assessment process resulting from the selection of people with known aMCI from memory clinics, and people with no cognitive impairment ("controls") from the community or from relatives of patients attending memory clinics. Twenty six studies were judged to be at an unclear risk of bias for the index or reference test interpretation since it was unclear the extent to which the tests were interpreted blindly. Ten of the studies were judged to be at high risk of bias in the flow and timing domain since patients and controls were not assessed with the same reference standard. Most studies (n=27) didn't report the time period between the index test and the reference standard and were therefore judged as unclear on this aspect.

[insert Figure 3 here]

Discussion

There is increasing interest in detecting people with aMCI as a potentially more timely point for treatment before the neuropathology has become more fully established with consequent dementia. The practical difficulty is that the diagnostic criteria for aMCI (the Petersen criteria (Petersen, 2004)) are resource intense to apply in routine care. Brief cognitive tests have therefore been investigated as a more practical first step in providing a quick indication of a person's cognitive state. The idea is not that these brief tests would replace the standard diagnostic criteria but that they could be used to quickly identify people who may have aMCI

and should be referred for further cognitive assessment. However, for them to be of use in identifying aMCI, a critical issue is to understand the diagnostic accuracy of the candidate tests. We therefore conducted a systematic search of the literature for studies that had reported evidence on the DTA of brief cognitive tests for aMCI. To ensure their applicability for clinical settings, and for potential community screening for case ascertainment in research studies, only those tools characterised as simple (not requiring specialist input or equipment), and quick (less than 15 minutes to administer) were included. Evidence for 42 cognitive screening tools that met these criteria was found.

The AVLT-SR was the most accurate single task index test with a high sensitivity (97%), high specificity (73%) and high overall diagnostic accuracy (AUC = 0.94). Other verbal learning tasks (CERAD-WLDR and HVLT-LE), as well as FBMS, which involves word recall, also exhibited high sensitivity for aMCI (83-84%). The high accuracy of these word recall tests is perhaps unsurprising since they assess episodic memory, a feature known to be impaired in aMCI and early AD (Petersen et al., 1999) and thought to be the result of early pathological changes in the medial temporal lobe (Braak and Braak, 1998).

Although episodic memory impairment is an important distinguishing feature of people with aMCI, studies have shown that non-memory cognitive impairments such as attention, processing speed, semantic fluency, executive function and visuospatial processing are also frequently detected in patients with aMCI (Economou et al., 2007). In this situation, the patient would be classified as multi-domain aMCI, that is, having cognitive impairment in memory and other non-memory domains. It has been reported that this form of aMCI may be more common than single-domain aMCI. For example, in a study by Alladi et al 2006, it was found that only 25 out of 90 patients with MCI had single-domain aMCI, and that deficits in

both semantic memory and attention were more common (Alladi et al., 2006). Another study by Diniz et al 2008 reported a higher proportion of their patients had multi-domain aMCI compared with single domain (59% vs. 29%) (Diniz et al., 2008a) again supporting this idea. In addition, multi-domain aMCI may be more likely to progress to dementia than single domain aMCI (for review see (Hughes et al., 2011)).

For these reasons assessment for impairment in multiple cognitive domains might be important in the identification of people with aMCI. Some single task cognitive but non-memory tests have shown promising results in the reviewed literature. For example, the AQT-CF which assesses perceptual speed and attention has a sensitivity of 84%. Nonetheless, multi-task tests that assess several cognitive domains provide the potential for a more comprehensive assessment. Of the multi-task tests identified in this systematic review, all provide an assessment of memory but in combination with various other cognitive domains. The MMSE was the most frequently reported multi-task index test. However, the reported sensitivities were generally unsatisfactory in comparison to the other multi-task tests. The MoCA was the next most frequently reported multi-task test. This test provides an assessment of five cognitive domains and takes 10-15 minutes to administer. Although a sensitivity to detect aMCI of 100% has been reported in association with a test score cut-off value of less than 24, for the most widely reported score cut-off value of 23/24 (four studies), the combined sensitivity was 86%. The M@T, which assesses episodic memory, semantic memory and orientation, and takes just 5-10 minutes to administer, also exhibited a very high sensitivity for aMCI (96%).

An important factor to consider when selecting a cognitive test is how its performance is influenced by demographic factors, such as age, gender and education level. Many of the

included studies controlled for such factors by ensuring that there were no significant differences between aMCI and control groups in these demographic characteristics. Other studies demonstrated a clear influence of education level on performance of the index test and consequently reported education-level dependent cut-off scores. In particular, Ladeira et al 2009 reported education-level dependent cut-off scores for MMSE, VFT-Animals and CDT-Sunderland, as did Zhao et al 2012 for AVLT-SR. The MoCA has also been shown to be influenced by education level, and it's developers (Nasreddine et al., 2005) recommend that one point is added to the score if the individual has 12 years or fewer of formal education.

Another aim of this review was to identify evidence for the validity of brief cognitive tests for predicting future dementia in those with aMCI. Only two longitudinal studies were identified that investigated this issue. Of the 12 cognitive tests investigated, the FCSRT, which assesses free and cued item recall, had the highest sensitivity (80%) for identifying people with aMCI who progressed to dementia at three year follow-up. Clearly more longitudinal studies are needed to support these findings and to extend this aspect of validity to other cognitive tests.

Test-retest reliability is another important property of cognitive testing. Reliability data was reported for seven of the included brief cognitive tests. The reliability of the MoCA was reported in eight studies with fairly high to high reliability reported (ICC = 0.75-0.92). High test-retest reliability was also reported for the AQT-CF (ICC = 0.88) whereas the ICCs for the MMSE tended to be lower (0.67-0.76).

The methodological quality of the included studies was also assessed in this review and only one study (McLennan et al., 2011) had a "low risk of bias" in all four assessed domains. This study assessed the validity of MoCA for detecting aMCI patients recruited from hospital

cardiovascular outpatient clinics and, although the MoCA detected all three patients with aMCI, it exhibited a low specificity of 50%. All other studies were judged to be at a high or unclear risk of bias and this therefore limits the confidence with which interpretations from the studies can be made. A large proportion of the studies were judged to be at a high risk of bias in the patient selection domain since they were at risk of unblinding the patient assessment process by recruiting patients with known aMCI from memory clinics and participants without cognitive impairment ("controls") from the community or via relatives of the patients. It has been reported that studies such as these may exaggerate diagnostic accuracy (Lijmer et al., 1999, Whiting et al., 2004). Another area that was assessed as high risk of bias for several studies was patient flow, where studies did not use the same reference standard for all participants. Improved study design should be a feature of future studies in this area.

Strengths of the review

This review used systematic method and followed standard guidance to provide a comprehensive summary of the literature on the diagnostic test accuracy of brief cognitive tests for aMCI. Two independent reviewers screened all potential studies for inclusion and extracted data, reducing potential risk of bias in study selection or errors in data extraction. All included studies were assessed for their methodological quality using a standardised tool (QUADAS-2). Finally, by ensuring that only those studies that used the Petersen criteria as the reference standard were included, the samples reported can be considered to be relatively homogeneous and comparable.

Weaknesses of the review

It is important to note that the prevalence of aMCI in most of the included studies was greater than that expected in the general population (reported to be 14-18% for individuals aged 70 years and older (Petersen et al., 2009)) and therefore calculated estimates of PPV/NPV from these studies are inflated and unlikely to be generalizable to older people in community settings. Also, some studies reported multiple thresholds and, in these cases, the optimal threshold reported by the author was chosen. This may have led to an overestimation of diagnostic accuracy (Leeflang et al., 2008), particularly in the summary ROC curves, where two included studies for MMSE and five included studies for MoCA reported multiple thresholds.

Conclusion

An ideal cognitive test for detecting people with aMCI would be one with high parameter values for DTA, predictive validity and test-retest reliability in the context of a well-designed experimental study. Of the 42 brief cognitive tests identified in this review, the MoCA was identified as the most comprehensively investigated test. The MoCA has a high sensitivity and high test-retest reliability but its predictive validity has yet to be investigated. Other brief cognitive tests, such as those that assess word recall (AVLT-SR, CERAD-WLDR, HVLT-LE and FBMS), and multi-task tests that assess several cognitive domains (such as M@T), have also been found to exhibit high sensitivities and reasonable specificities. However, lack of evidence on the predictive validity of these tests, and concerns over the quality of the constituent studies, limit the confidence with which definitive recommendations can be made. Further studies validating the most promising cognitive tests to detect aMCI is warranted.

Acknowledgements

We thank Deirdre Andre for conducting the database searches, Sarah Smith for assisting with study selection, Faye Plummer for providing advice on the meta-analysis procedure and Jean-Francois Delvenne for providing guidance on the review. This review was supported by a Research for Patient Benefit grant awarded by the National Institute for Health Research.

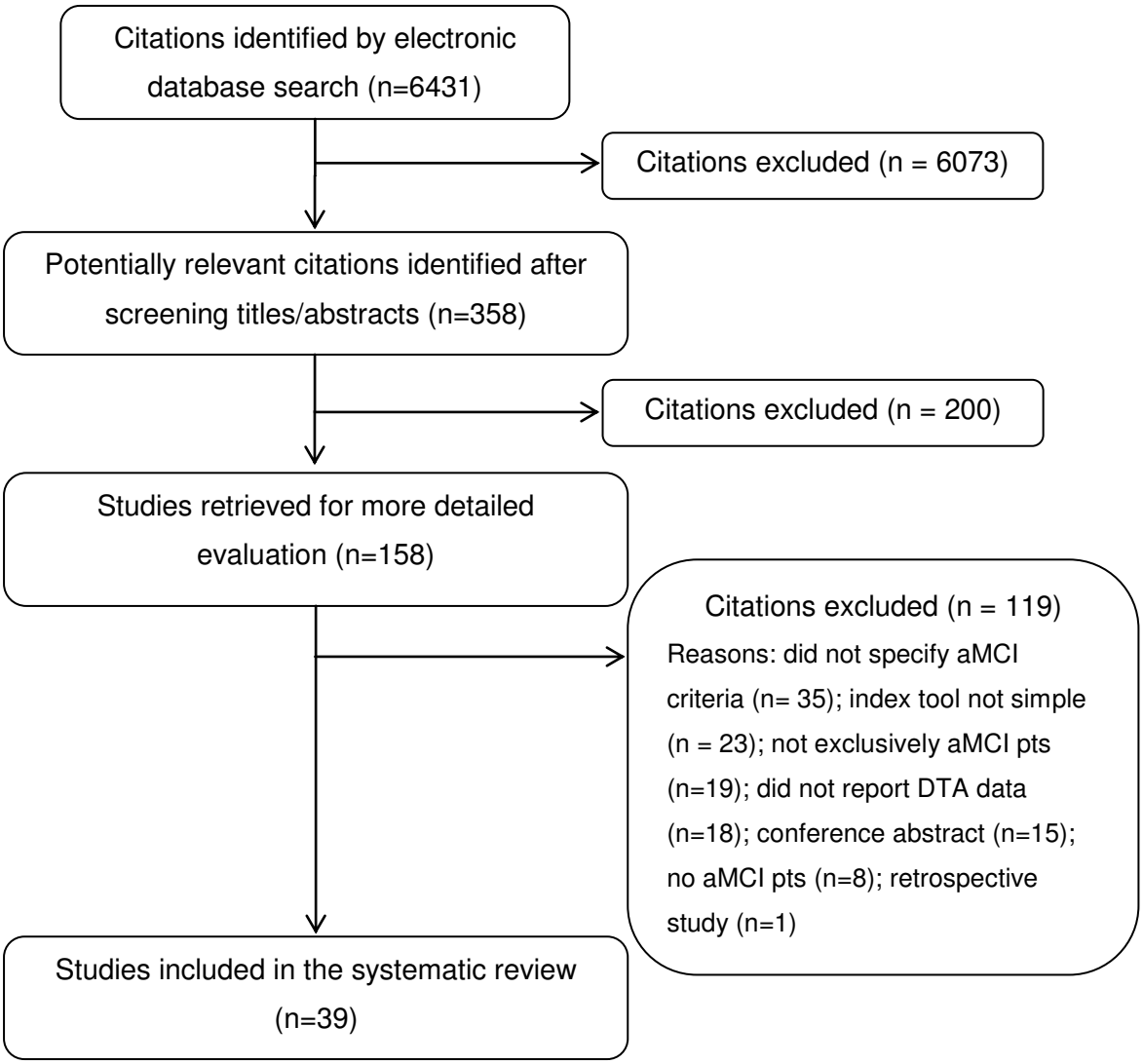

Figure 1: Study Selection Process Diagram (using the PRISMA guidelines)

Table 1: Characteristics of included studies

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
Cross-sectional DTA studies								
(Ahmed et al., 2012)	2012	UK	35	42.9	Community (OPTIMA cohort)	80.9 (7.2)	MoCA	Unclear
(Ahn et al., 2010)	2010	Korea	120	35.8	Memory clinic	NR	MMSE	High
(Alegret et al., 2009)	2009	Spain	88	50.0	Diagnostic unit of Fundacio ACE	76.5 (5.1)	15-Objects Test	High
(Cacho et al., 2010)	2010	Spain	87	24.1	Memory clinic	73.8 (5.0)	CDT-command; MMSE (alone & in combination)	High
(Chandler et al., 2005)	2005	USA	155	38.7	University Clinic for Alzheimer's and Related Diseases	72.8 (7.5)	VLT (CERAD-WLDR); MMSE	High
(Costa et al., 2012)	2012	Germany	130	23.1	Memory clinic	67.8 (8.1)	MoCA	High
(Dierckx et al., 2007)	2007	Belgium	92	43.5	Memory clinic & psychiatric hospital	75.0 (6.0)	MIS-plus & VAT	High
(Diniz et al., 2008b)	2008	Brazil	165	46.1	Memory clinic	72.3 (6.6)[SD] 70.2 (6.5)[MD]	MMSE	High
(Freitas et al., 2013)	2013	Portugal	270	33.3	Dementia clinic	70.5 (8.0)	MMSE; MoCA	High
(Fujiwara et al., 2010)	2010	Japan	66	45.5	Memory clinic	77.3 (6.3)	HDS-R; MoCA	High
(Gonzalez-Palau et al., 2013)	2013	Spain	241	54.8	Memory clinic, residential facilities, community centres	82.0 (9.2)	VLT (HVLT LE); MMSE	High
(Guo et al., 2012)	2012	China	508	61.2	Memory clinic	70.0 (9.1)[SD] 70.3 (8.8)[MD]	MES; MMSE	High

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
(Hanyu et al., 2009)	2009	Japan	63	49.2	Memory clinic	75.6 (5.1)	VFT-Animals	High
(Hanyu et al., 2011)	2011	Japan	80	57.5	Memory clinic	76.0 (6.6)	MMSE; TYM	High
(Karrasch et al., 2005)	2005	Finland	30	50.0	Neurologist referral	67.5 (9.2)	CDT-CERAD; Constructional Praxis-Savings; Naming-BNT-M; VFT-Animals; VLTs (CERAD-WLDR, -WLLE, -WLRE, WLSA); MMSE	High
(Kato et al., 2013)	2013	Japan	109	55.0	Hospital	76.1 (9.2)	CDT-command; MMSE	High
(Ladeira et al., 2009)	2009	Brazil	166	50.0	Memory clinic	70.3 (6.1)	CDT-Sunderland; VFT-Animals; MMSE (alone & in combination)	Unclear
(Lee et al., 2008)	2008	Korea	152	24.3	Hospital & community	71.3 (5.9)	MoCA	Unclear
(Loewenstein et al., 2009)	2009	USA	103	22.3	Centre for AD and memory disorders	79.7 (6.0)	FBMS	High
(Luis et al., 2009)	2009	USA	98	24.5	Memory clinic & community	78.9 (5.3)	MMSE; MoCA	Unclear
(McLennan et al., 2011)	2011	Australia	98	3.1	Cardiac & diabetic/ endocrine outpatient clinics	NR	MoCA	Low

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
(Muangpaisan et al., 2010)	2010	Thailand	107	72.0	Community (BLOSSOM cohort)	66.3 (7.9)	Digit Span (Forward & Backward); VFTs (Animals, Fruits, Letter Koh & Soh)	Unclear
(Nasreddine et al., 2005)	2005	Canada	184	51.1	Memory clinic	75.2 (6.3)	MMSE; MoCA	High
(Rahman and El Gaafary, 2009)	2009	Egypt	184	51.1	Community (geriatric clubs)	NR	MoCA	Unclear
(Rami et al., 2007)	2007	Spain	450	11.1	Memory-Alzheimer's Unit Hospital Clinic	76.6 (6.6)	M@T	High
(Rami et al., 2010)	2010	Spain	87	57.5	Memory clinic	76.6 (6.6)	M@T	Unclear
(Ravaglia et al., 2005)	2005	Italy	93	40.1	University Centre for Physiopathology of Aging	76.5 (7.1)	CDT (Sunderland, Wolf-Klein); MMSE (alone & in combination)	High
(Saka et al., 2006)	2006	Turkey	51	35.3	Dementia outpatient clinic	69.4 (8.3)	ECR (3rd free & total recall)	High
(Scheurich et al., 2005)	2005	Germany	20	65.0	Memory clinic	66.4 (9.7)	DemTect	Unclear
(Schrijnemaekers et al., 2006)	2006	UK	73	26.0	Community (Foresight Challenge study)	76.2 (9.4)	MMSE; VLT (HVLT LE)	High
(Smith et al., 2007)	2007	UK	35	65.7	Memory clinic	77.5 (7.8)	MMSE; MoCA	Unclear
(Takahashi et al., 2012)	2012	Japan	50	50.0	Medical centre for dementia	75.2 (5.4)	AQT-CF	Unclear
(Tsai et al., 2012)	2012	Taiwan	109	65.1	Memory clinic	79.2 (6.8)	MoCA	High

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
(Woodard et al., 2005)	2005	USA	179	10.1	General Internal Medicine & Geriatric clinics	75.9 (5.7)	VFT-Animals; VLTs (CERAD-WLDR, -WLREDI, -WLSA)	Unclear
(Yoshida et al., 2012)	2012	Japan	112	34.8	Memory clinic	71.4 (9.2)	MMSE	High
(Zhao et al., 2011)	2011	China	300	50.0	Hospital	70.7 (4.3)	MoCA	High
(Zhao et al., 2012)	2012	China	641^	50.7^	Memory clinic	74.1 (2.8)^	VLT (AVLT SR)	High
Longitudinal Studies								
(Ahmed et al., 2008)	2008	UK	18[A]	38.9[P]	Memory clinic	71.7 (6.8)[P] 71.3 (7.7)[NP]	Naming-GNT; TMT-Part B; VFT-Animals	High
(Sarazin et al., 2007)	2007	France	217[A]	27.2[P]	Memory clinic	74.8 (4.1)[P] 70.9 (5.4)[NP]	FCSRT (Total & Free Recall); Serial digit ordering; Stroop - inhibition; TMT (A & B); VFTs (Fruits & "S"); WAIS (Similarities & Digit Symbol)	High

KEY: *aMCI and cognitively normal participants only; ^70-79yrs age group only; [A]aMCI participants only; [MD]multi-domain aMCI; [NP]aMCI non-progressors; [P]aMCI progressors; [SD]single domain aMCI; AD = Alzheimer's disease; NR = not reported
NB: for index test abbreviations see Glossary

Table 2: Summary of diagnostic test accuracy results for single task cognitive tests for identifying aMCI

Index Test (units)	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
MEMORY TASKS									
ECR									
-3rd Free Recall	9	56	79	59	76	2.62	0.56	0.69	(Saka et al., 2006)
-Total Recall	42	50	91	75	77	5.50	0.55	0.63	(Saka et al., 2006)
FBMS	7[OR]	83	88	66	95	6.61	0.20	0.90	(Loewenstein et al., 2009)
Verbal Learning Task									
-AVLT SR	≤2[^]	97	73	79	95	3.55	0.05	0.94	(Zhao et al., 2012)
-CERAD WLDR	6	27	100	100	57	-	0.73	NR	(Karrasch et al., 2005)
	6.5	82	63	58	85	2.22	0.29	0.82	(Chandler et al., 2005)
	<7	83	60	19	97	2.10	0.28	0.76	(Woodard et al., 2005)
-CERAD WLLE	20[OS]	73	80	79	75	3.67	0.33	NR	(Karrasch et al., 2005)
-CERAD WLREDI	<10	94	35	14	98	1.45	0.16	0.73	(Woodard et al., 2005)
-CERAD WLRE (%)	92[OS]	47	93	88	64	7.00	0.57	NR	(Karrasch et al., 2005)
-CERAD WLSA (%)	<80	89	55	18	98	1.99	0.20	0.77	(Woodard et al., 2005)
	80	33	67	50	50	1.00	1.00	NR	(Karrasch et al., 2005)
-HVLT LE	≤15[OR]	83	65	74	76	2.39	0.26	0.84	(Gonzalez-Palau et al., 2013)
	24.5	84	80	59	93	4.13	0.20	NR	(Schrijnemaekers et al., 2006)
Constructional Praxis-Savings (%)	60	33	67	50	50	1.00	1.00	NR	(Karrasch et al., 2005)
Digit Span									
-Forward	12	64	70	84	43	2.12	0.52	0.71	(Muangpaisan et al., 2010)
-Backward	4	77	57	82	49	1.77	0.41	0.73	(Muangpaisan et al., 2010)
NON-MEMORY TASKS									
Naming-BNT-M	11	13	100	100	53	-	0.87	NR	(Karrasch et al., 2005)
Verbal Fluency Task									
-Animals	ELD[$]	27	95	85	56	5.50	0.77	0.61	(Ladeira et al., 2009)
	14	81	69	71	79	2.58	0.28	NR	(Hanyu et al., 2009)
		83	43	79	50	1.47	0.39	0.63	(Muangpaisan et al., 2010)
	15	27	100	100	57	-	0.73	NR	(Karrasch et al., 2005)
	<20	72	55	15	95	1.62	0.50	0.69	(Woodard et al., 2005)
-Fruits	15	68	63	83	43	1.84	0.51	0.69	(Muangpaisan et al., 2010)

Index Test (units)	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
-Letter Koh	9	50	73	83	36	1.88	0.68	0.66	(Muangpaisan et al., 2010)
-Letter Soh	7	81	57	83	53	1.86	0.34	0.71	(Muangpaisan et al., 2010)
15-Objects Test	12	64	86	82	70	4.67	0.42	0.85	(Alegret et al., 2009)
AQT-CF (seconds)	72/73	84	76	78	83	3.50	0.21	0.88	(Takahashi et al., 2012)
Clock Drawing Test									
-CERAD	5	7	87	33	48	0.50	1.08	NR	(Karrasch et al., 2005)
-Command	8/9	76	70	44	90	2.51	0.34	0.78	(Cacho et al., 2010)
		50	88	83	59	4.08	0.57	0.72	(Kato et al., 2013)
-Sunderland	ELD*	30	88	71	56	2.50	0.79	0.59	(Ladeira et al., 2009)
	≤5	26	85	56	63	1.81	0.86	NR	(Ravaglia et al., 2005)
-Wolf Klein	≤6	21	89	57	62	1.93	0.89	NR	(Ravaglia et al., 2005)

KEY: ^70-79 years age group; AUC = Area Under Receiver Operating Characteristic (ROC) curve; ELD* = education-level dependent (0-8 years of education <6; >8 years of education <8); ELD$ = education-level dependent (illiterate <10, 1+ year of education <14); LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio; NPV = Negative Predictive Value; NR = Not Reported; PPV = Positive Predictive Value; ORindicates >1 threshold was reported in the study but only author-reported optimal threshold was extracted; OSindicates >1 threshold was reported in the study but only threshold with maximum sensitivity was extracted

NB: for index test abbreviations see Glossary

Table 3: Summary of diagnostic test accuracy results for multi-task cognitive tests for identifying aMCI

Index Test	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
DemTect	≤13	85	86	92	75	5.92	0.18	0.92	(Scheurich et al., 2005)
HDS-R	28/29	87	61	65	85	2.23	0.22	0.86	(Fujiwara et al., 2010)
M@T	37[OR]	96	79	36	99	4.57	0.05	0.93	(Rami et al., 2007)
		96	70	81	93	3.23	0.06	0.88	(Rami et al., 2010)
MES	≤72	88	91	91	88	10.2	0.13	0.96	(Guo et al., 2012)[MD]
	≤75	79	83	73	87	4.60	0.25	0.89	(Guo et al., 2012)[SD]
MMSE	<24	29	87	61	64	2.27	0.81	NR	(Ravaglia et al., 2005)
	24v25	52	95	79	86	11.5	0.50	0.82	(Cacho et al., 2010)
	25	13	93	67	52	2.00	0.93	NR	(Karrasch et al., 2005)
	<26	18	100	100	54	-	0.82	NR	(Nasreddine et al., 2005)
	≤26[OR]	76	69	75	71	2.43	0.34	0.76	(Gonzalez-Palau et al., 2013)
	26	17	100	100	39	-	0.83	NR	(Smith et al., 2007)
	26/27	63	96	95	68	15.5	0.38	0.84	(Kato et al., 2013)
		41	99	96	76	41.0	0.60	NR	(Yoshida et al., 2012)
	≤27	68	61	50	77	1.77	0.52	0.67	(Guo et al., 2012)[SD]
		68	70	69	69	2.26	0.46	0.72	(Guo et al., 2012)[MD]
		58	84	54	86	3.60	0.50	0.76	(Luis et al., 2009)[OR]
	27/28	72	60	33	88	1.78	0.47	0.72	(Diniz et al., 2008b)*
		71	61	51	78	1.80	0.48	0.73	(Diniz et al., 2008b)[$]
		70	68	74	62	2.15	0.45	0.73	(Hanyu et al., 2011)
	28.5	60	70	53	76	2.02	0.56	0.72	(Ahn et al., 2010)
		67	61	52	74	1.71	0.55	0.69	(Chandler et al., 2005)
		74	69	45	88	2.34	0.38	NR	(Schrijnemaekers et al., 2006)
	<29	67	72	71	68	2.40	0.46	0.75	(Freitas et al., 2013)
	ELD	54	71	65	61	1.88	0.64	0.63	(Ladeira et al., 2009)

Index Test	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
MoCA	<22	81	77	78	80	3.48	0.25	0.86	(Freitas et al., 2013)
	22/23[OR]	89	84	65	96	5.70	0.13	0.94	(Lee et al., 2008)
	23[OR]	96	95	85	99	17.7	0.04	0.97	(Luis et al., 2009)
	23.5	88	65	67	87	2.50	0.19	0.89	(Ahmed et al., 2012)
	23/24	86	86	-	-	5.93	0.17	-	(Fujiwara et al., 2010, Lee et al., 2008, Tsai et al., 2012, Zhao et al., 2011)[&]
	<24[OR]	100	50	6	100	2.00	0.00	NR	(McLennan et al., 2011)
	25/26[OR]	93	89	88	94	8.40	0.08	0.95	(Fujiwara et al., 2010)
	<26	90	87	88	90	6.77	0.11	NR	(Nasreddine et al., 2005)
		93	86	87	92	6.41	0.09	NR	(Rahman and El Gaafary, 2009)
	≤26	93	62	42	97	2.46	0.11	0.85	(Costa et al., 2012)
	26	83	50	76	60	1.65	0.35	NR	(Smith et al., 2007)
TYM	44/45	76	74	80	69	2.87	0.33	0.86	(Hanyu et al., 2011)

KEY: [&] meta-analysis of 4 studies; [MD]multi-domain aMCI; [SD]single domain aMCI; AUC = Area Under Receiver Operating Characteristic (ROC) curve; ELD = education-level dependent (illiterate: <20, 1-4 years education: <25, 4-8 years education: <26, 9+ years education: <28); LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio; NPV = Negative Predictive Value; NR = Not Reported; PPV = Positive Predictive Value; [OR]indicates >1 threshold was reported in the study but only author-reported optimal threshold was extracted; [OS]indicates >1 threshold was reported in the study but only threshold with maximum sensitivity was extracted

NB: for index test abbreviations see Glossary

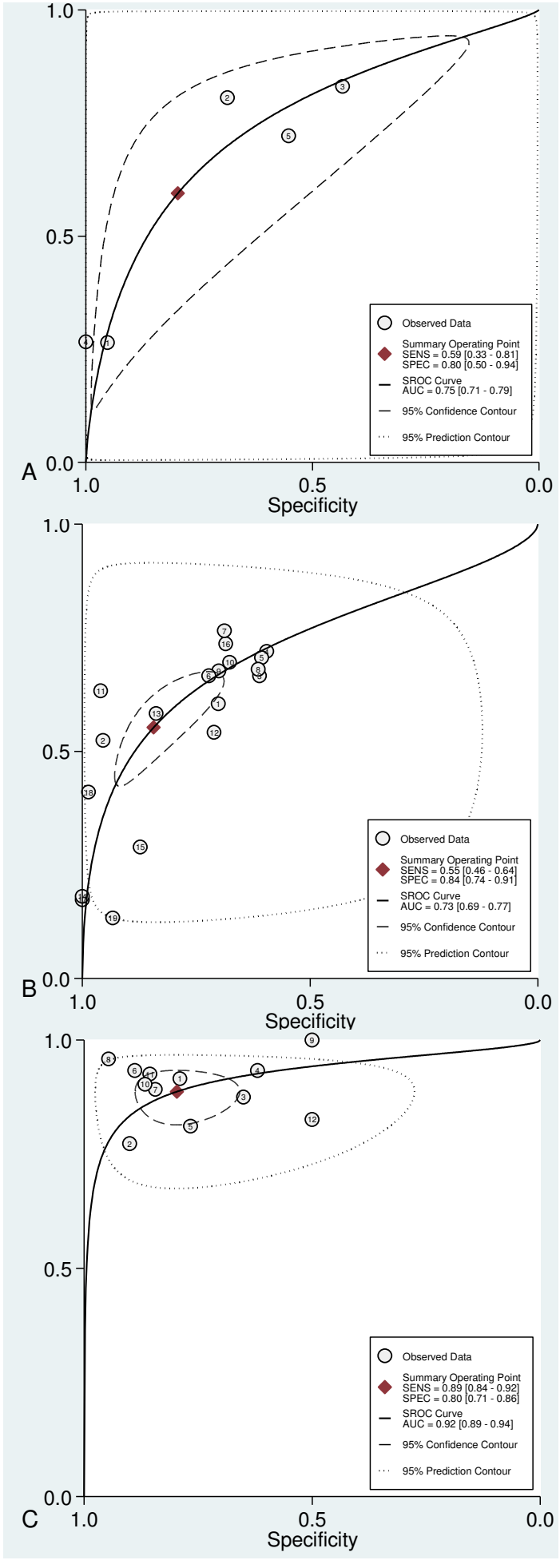

Figure 2: Summary receiver operating characteristic plots for studies using (A) VFT-Animals, (B) MMSE and (C) MoCA

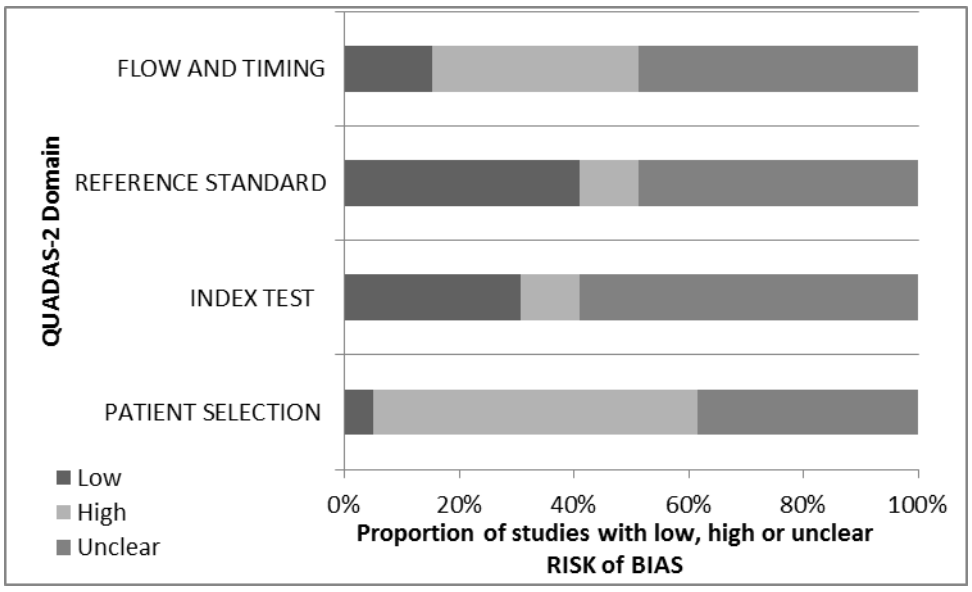

Figure 3: Summary of risk of bias judgements across all studies

References

AHMED, S., DE JAGER, C. & WILCOCK, G. 2012. A comparison of screening tools for the assessment of mild cognitive impairment: preliminary findings. *Neurocase,* 18**,** 336-51.

AHMED, S., MITCHELL, J., ARNOLD, R., NESTOR, P. J. & HODGES, J. R. 2008. Predicting rapid clinical progression in amnestic mild cognitive impairment. *Dementia & Geriatric Cognitive Disorders,* 25**,** 170-7.

AHN, H.-J., CHIN, J., PARK, A., LEE, B. H., SUH, M. K., SEO, S. W. & NA, D. L. 2010. Seoul Neuropsychological Screening Battery-dementia version (SNSB-D): a useful tool for assessing and monitoring cognitive impairments in dementia patients. *Journal of Korean Medical Science,* 25**,** 1071-6.

ALEGRET, M., BOADA-ROVIRA, M., VINYES-JUNQUE, G., VALERO, S., ESPINOSA, A., HERNANDEZ, I., MODINOS, G., ROSENDE-ROCA, M., MAULEON, A., BECKER, J. T. & TARRAGA, L. 2009. Detection of visuoperceptual deficits in preclinical and mild Alzheimer's disease. *Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society,* 31**,** 860-7.

ALLADI, S., ARNOLD, R., MITCHELL, J., NESTOR, P. J. & HODGES, J. R. 2006. Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychol Med,* 36**,** 507-15.

BRAAK, H. & BRAAK, E. 1998. Evolution of neuronal changes in the course of Alzheimer's disease. *J Neural Transm Suppl,* 53**,** 127-40.

CACHO, J., BENITO-LEON, J., GARCIA-GARCIA, R., FERNANDEZ-CALVO, B., VICENTE-VILLARDON, J. L. & MITCHELL, A. J. 2010. Does the combination of the MMSE and clock drawing test (mini-clock) improve the detection of mild Alzheimer's disease and mild cognitive impairment? *Journal of Alzheimer's Disease,* 22**,** 889-96.

CHANDLER, M. J., LACRITZ, L. H., HYNAN, L. S., BARNARD, H. D., ALLEN, G., DESCHNER, M., WEINER, M. F. & CULLUM, C. M. 2005. A total score for the CERAD neuropsychological battery. *Neurology,* 65**,** 102-6.

COSTA, A. S., FIMM, B., FRIESEN, P., SOUNDJOCK, H., ROTTSCHY, C., GROSS, T., EITNER, F., REICH, A., SCHULZ, J. B., NASREDDINE, Z. S. & REETZ, K. 2012. Alternate-form reliability of the Montreal cognitive assessment screening test in a clinical setting. *Dementia & Geriatric Cognitive Disorders,* 33**,** 379-84.

DEEKS, J., BOSSUYT, P. & GATSONIS, C. 2010. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*.

DIERCKX, E., ENGELBORGHS, S., DE RAEDT, R., DE DEYN, P. P. & PONJAERT-KRISTOFFERSEN, I. 2007. Differentiation between mild cognitive impairment, Alzheimer's disease and depression by means of cued recall. *Psychological Medicine,* 37**,** 747-55.

DINIZ, B. S., NUNES, P. V., YASSUDA, M. S., PEREIRA, F. S., FLAKS, M. K., VIOLA, L. F., RADANOVIC, M., ABREU, I. D., BORELLI, D. T., GATTAZ, W. F. & FORLENZA, O. V. 2008a. Mild cognitive impairment: cognitive screening or neuropsychological assessment? *Rev Bras Psiquiatr,* 30**,** 316-21.

DINIZ, B. S., NUNES, P. V., YASSUDA, M. S., PEREIRA, F. S., FLAKS, M. K., VIOLA, L. F., RADANOVIC, M., ABREU, I. D. D., BORELLI, D. T., GATTAZ, W. F. & FORLENZA, O. V. 2008b. Mild cognitive impairment: cognitive screening or neuropsychological assessment? *Revista Brasileira de Psiquiatria,* 30**,** 316-21.

DWAMENA, B. A., SYLVESTER, R. & CARLOS, R. C. 2010. Midas: meta-analysis of diagnostic accuracy studies. Available: http://fmwww.bc.edu/repec/bocode/m/midas.pdf.

ECONOMOU, A., PAPAGEORGIOU, S. G., KARAGEORGIOU, C. & VASSILOPOULOS, D. 2007. Nonepisodic memory deficits in amnestic MCI. *Cogn Behav Neurol,* 20**,** 99-106.

FERRI, C. P., PRINCE, M., BRAYNE, C., BRODATY, H., FRATIGLIONI, L., GANGULI, M., HALL, K., HASEGAWA, K., HENDRIE, H., HUANG, Y., JORM, A., MATHERS, C., MENEZES, P. R., RIMMER, E. & SCAZUFCA, M. 2005. Global prevalence of dementia: a Delphi consensus study. *Lancet,* 366**,** 2112-7.

FREITAS, S., SIMOES, M. R., ALVES, L. & SANTANA, I. 2013. Montreal cognitive assessment: Validation study for mild cognitive impairment and alzheimer disease. *Alzheimer Disease and Associated Disorders,* 27**,** 37-43.

FUJIWARA, Y., SUZUKI, H., YASUNAGA, M., SUGIYAMA, M., IJUIN, M., SAKUMA, N., INAGAKI, H., IWASA, H., URA, C., YATOMI, N., ISHII, K., TOKUMARU, A. M., HOMMA, A., NASREDDINE, Z. & SHINKAI, S. 2010. Brief screening tool for mild cognitive impairment in older Japanese: validation of the Japanese version of the Montreal Cognitive Assessment. *Geriatrics & gerontology international,* 10**,** 225-32.

GAUTHIER, S., REISBERG, B., ZAUDIG, M., PETERSEN, R. C., RITCHIE, K., BROICH, K., BELLEVILLE, S., BRODATY, H., BENNETT, D., CHERTKOW, H., CUMMINGS, J. L., DE LEON, M., FELDMAN, H., GANGULI, M., HAMPEL, H., SCHELTENS, P., TIERNEY, M. C., WHITEHOUSE, P. & WINBLAD, B. 2006. Mild cognitive impairment. *Lancet,* 367**,** 1262-70.

GONZALEZ-PALAU, F., FRANCO, M., JIMENEZ, F., PARRA, E., BERNATE, M. & SOLIS, A. 2013. Clinical utility of the Hopkins Verbal Test-Revised for detecting Alzheimer's disease and mild cognitive impairment in Spanish population. *Archives of Clinical Neuropsychology,* 28**,** 245-253.

GUO, Q.-H., ZHOU, B., ZHAO, Q.-H., WANG, B. & HONG, Z. 2012. Memory and Executive Screening (MES): a brief cognitive test for detecting mild cognitive impairment. *BMC Neurology,* 12**,** 119.

HANYU, H., KUME, K., TAKADA, Y., ONUMA, T. & IWAMOTO, T. 2009. The 1-minute Mental Status Examination in the memory clinic. *Journal of the American Geriatrics Society,* 57**,** 1130-1131.

HANYU, H., MAEZONO, M., SAKURAI, H., KUME, K., KANETAKA, H. & IWAMOTO, T. 2011. Japanese version of the Test Your Memory as a screening test in a Japanese memory clinic. *Psychiatry Research,* 190**,** 145-8.

HARBORD, R. M., DEEKS, J. J., EGGER, M. S., WHITING, P. & STERNE, J. A. 2007. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics,* 8**,** 239-251.

HUGHES, T. F., SNITZ, B. E. & GANGULI, M. 2011. Should mild cognitive impairment be subtyped? *Curr Opin Psychiatry,* 24**,** 237-42.

KARRASCH, M., SINERVA, E., GRONHOLM, P., RINNE, J. & LAINE, M. 2005. CERAD test performances in amnestic mild cognitive impairment and Alzheimer's disease. *Acta Neurologica Scandinavica,* 111**,** 172-9.

KATO, Y., NARUMOTO, J., MATSUOKA, T., OKAMURA, A., KOUMI, H., KISHIKAWA, Y., TERASHIMA, S. & FUKUI, K. 2013. Diagnostic performance of a combination of mini-mental state examination and clock drawing test in detecting Alzheimer's disease. *Neuropsychiatric Disease and Treatment,* 9**,** 581-586.

LADEIRA, R. B., DINIZ, B. S., NUNES, P. V. & FORLENZA, O. V. 2009. Combining cognitive screening tests for the evaluation of mild cognitive impairment in the elderly. *Clinics (Sao Paulo, Brazil),* 64**,** 967-73.

LEE, J.-Y., DONG WOO, L., CHO, S.-J., NA, D. L., HONG JIN, J., KIM, S.-K., YOU RA, L., YOUN, J.-H., KWON, M., LEE, J.-H. & MAENG JE, C. 2008. Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry & Neurology,* 21**,** 104-10.

LEEFLANG, M. M., MOONS, K. G., REITSMA, J. B. & ZWINDERMAN, A. H. 2008. Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. *Clinical Chemistry,* 54**,** 729-737.

LIJMER, J. G., MOL, B. W., HEISTERKAMP, S., BONSEL, G. J., PRINS, M. H., VAN DER MEULEN, J. H. & BOSSUYT, P. M. 1999. Empirical evidence of design-related bias in studies of diagnostic tests. *Jama,* 282**,** 1061-6.

LOEWENSTEIN, D. A., ACEVEDO, A., AGRON, J., OWNBY, R., BARKER, W., STRAUMAN, S. & DUARA, R. 2009. The utility of a brief memory screen in the diagnosis of mild memory impairment in the elderly: preliminary study. *American Journal of Geriatric Psychiatry,* 17**,** 437-40.

LONIE, J. A., TIERNEY, K. M. & EBMEIER, K. P. 2009. Screening for mild cognitive impairment: a systematic review. *Int J Geriatr Psychiatry,* 24**,** 902-15.

LUIS, C. A., KEEGAN, A. P. & MULLAN, M. 2009. Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *International Journal of Geriatric Psychiatry,* 24**,** 197-201.

MCLENNAN, S. N., MATHIAS, J. L., BRENNAN, L. C. & STEWART, S. 2011. Validity of the montreal cognitive assessment (MoCA) as a screening test for mild cognitive impairment (MCI) in a cardiovascular population. *Journal of Geriatric Psychiatry & Neurology,* 24**,** 33-8.

MOHER, D., LIBERATI, A., TETZLAFF, J., ALTMAN, D. G. & GROUP, T. P. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine,* 6**,** 1-6.

MUANGPAISAN, W., INTALAPAPORN, S. & ASSANTACHAI, P. 2010. Digit span and verbal fluency tests in patients with mild cognitive impairment and normal subjects in Thai-community. *Journal of the Medical Association of Thailand,* 93**,** 224-30.

NASREDDINE, Z. S., PHILLIPS, N. A., BEDIRIAN, V., CHARBONNEAU, S., WHITEHEAD, V., COLLIN, I., CUMMINGS, J. L. & CHERTKOW, H. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society,* 53**,** 695-9.

PETERSEN, R. C. 2004. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine,* 256**,** 183-94.

PETERSEN, R. C., DOODY, R., KURZ, A., MOHS, R. C., MORRIS, J. C., RABINS, P. V., RITCHIE, K., ROSSOR, M., THAL, L. & WINBLAD, B. 2001. Current concepts in mild cognitive impairment. *Arch Neurol,* 58**,** 1985-92.

PETERSEN, R. C., ROBERTS, R. O., KNOPMAN, D. S., BOEVE, B. F., GEDA, Y. E., IVNIK, R. J., SMITH, G. E. & JACK, C. R., JR. 2009. Mild cognitive impairment: ten years later. *Arch Neurol,* 66**,** 1447-55.

PETERSEN, R. C., SMITH, G. E., WARING, S. C., IVNIK, R. J., TANGALOS, E. G. & KOKMEN, E. 1999. Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology,* 56**,** 303-8.

RAHMAN, T. T. A. & EL GAAFARY, M. M. 2009. Montreal Cognitive Assessment Arabic version: reliability and validity prevalence of mild cognitive impairment among elderly attending geriatric clubs in Cairo. *Geriatrics & gerontology international,* 9**,** 54-61.

RAMI, L., BOSCH, B., SANCHEZ-VALLE, R. & MOLINUEVO, J. L. 2010. The memory alteration test (M@T) discriminates between subjective memory complaints, mild cognitive impairment and Alzheimer's disease. *Archives of Gerontology & Geriatrics,* 50**,** 171-4.

RAMI, L., MOLINUEVO, J. L., SANCHEZ-VALLE, R., BOSCH, B. & VILLAR, A. 2007. Screening for amnestic mild cognitive impairment and early Alzheimer's disease with M@T (Memory Alteration Test) in the primary care population. *International Journal of Geriatric Psychiatry,* 22**,** 294-304.

RAVAGLIA, G., FORTI, P., MAIOLI, F., SERVADEI, L., MARTELLI, M., BRUNETTI, N., BASTAGLI, L. & MARIANI, E. 2005. Screening for mild cognitive impairment in elderly ambulatory patients with cognitive complaints. *Aging-Clinical & Experimental Research,* 17**,** 374-9.

REITSMA, J. B., GLAS, A. S., RUTJES, A. W., SCHOLTEN, R. J. P. M., BOSSUYT, P. & ZWINDERMAN, A. H. 2005. Bivariate analysis of sensitivity and specifcity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology,* 58**,** 982-990.

ROBERTS, R. O., GEDA, Y. E., KNOPMAN, D. S., CHA, R. H., PANKRATZ, V. S., BOEVE, B. F., TANGALOS, E. G., IVNIK, R. J., ROCCA, W. A. & PETERSEN, R. C. 2012. The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging. *Neurology,* 78**,** 342-51.

SAKA, E., MIHCI, E., TOPCUOGLU, M. A. & BALKAN, S. 2006. Enhanced cued recall has a high utility as a screening test in the diagnosis of Alzheimer's disease and mild cognitive impairment in Turkish people. *Archives of Clinical Neuropsychology,* 21**,** 745-51.

SARAZIN, M., BERR, C., DE ROTROU, J., FABRIGOULE, C., PASQUIER, F., LEGRAIN, S., MICHEL, B., PUEL, M., VOLTEAU, M., TOUCHON, J., VERNY, M. & DUBOIS, B. 2007. Amnestic syndrome of the medial temporal type identifies prodromal AD: a longitudinal study.[Erratum appears in Neurology. 2008 May 20;70(21):2016]. *Neurology,* 69**,** 1859-67.

SCHEURICH, A., MULLER, M. J., SIESSMEIER, T., BARTENSTEIN, P., SCHMIDT, L. G. & FELLGIEBEL, A. 2005. Validating the DemTect with 18-fluoro-2-deoxy-glucose positron emission tomography as a sensitive neuropsychological screening test for early alzheimer disease in patients of a memory clinic. *Dementia & Geriatric Cognitive Disorders,* 20**,** 271-7.

SCHRIJNEMAEKERS, A. M. C., DE JAGER, C. A., HOGERVORST, E. & BUDGE, M. M. 2006. Cases with mild cognitive impairment and Alzheimer's disease fail to benefit from repeated exposure to episodic memory tests as compared with controls. *Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society,* 28**,** 438-55.

SMITH, T., GILDEH, N. & HOLMES, C. 2007. The Montreal Cognitive Assessment: validity and utility in a memory clinic setting. *Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie,* 52**,** 329-32.

TAKAHASHI, F., AWATA, S., SAKUMA, N., INAGAKI, H. & IJUIN, M. 2012. Reliability and validity of A Quick Test of Cognitive Speed for detecting early-stage dementia in elderly Japanese. *Psychogeriatrics:The Official Journal of the Japanese Psychogeriatric Society,* 12**,** 75-82.

TSAI, C. F., LEE, W. J., WANG, S. J., SHIA, B. C., NASREDDINE, Z. & FUH, J. L. 2012. 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,* 24**,** 651-658.

WHITING, P., RUTJES, A. W., REITSMA, J. B., GLAS, A. S., BOSSUYT, P. M. & KLEIJNEN, J. 2004. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med,* 140**,** 189-202.

WHITING, P. F., RUTJES, A. W., WESTWOOD, M. E., MALLETT, S., DEEKS, J. J., REITSMA, J. B., LEEFLANG, M. M., STERNE, J. A. & BOSSUYT, P. M. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med,* 155**,** 529-36.

WOODARD, J. L., DORSETT, E. S. W., COOPER, J. G., HERMANN, B. P. & SAGER, M. A. 2005. Development of a brief cognitive screen for mild cognitive impairment and neurocognitive disorder. *Aging Neuropsychology and Cognition,* 12**,** 299-315.

YOSHIDA, H., TERADA, S., HONDA, H., KISHIMOTO, Y., TAKEDA, N., OSHIMA, E., HIRAYAMA, K., YOKOTA, O. & UCHITOMI, Y. 2012. Validation of the revised Addenbrooke's Cognitive Examination (ACE-R) for detecting mild cognitive impairment and dementia in a Japanese population. *Int Psychogeriatr,* 24**,** 28-37.

ZHAO, Q., LV, Y., ZHOU, Y., HONG, Z. & GUO, Q. 2012. Short-Term Delayed Recall of Auditory Verbal Learning Test Is Equivalent to Long-Term Delayed Recall for Identifying Amnestic Mild Cognitive Impairment. *PLoS ONE [Electronic Resource],* 7.

ZHAO, S., GUO, C., WANG, M., CHEN, W., WU, Y., TANG, W. & ZHAO, Y. 2011. A clinical memory battery for screening for amnestic mild cognitive impairment in an elderly chinese population. *Journal of Clinical Neuroscience,* 18**,** 774-9.