Cross-Cultural applicability of the Montreal Cognitive Assessment (MoCA): a systematic review

Abstract

The MoCA is widely used to screen for mild cognitive impairment. While there are many available versions, the cross cultural validity of the assessment has not been explored sufficiently.

We aimed to interrogate the validity of the MoCA in a cross-cultural context: in differentiating mild cognitive impairment (MCI) from normal controls (NC); and identifying cut-offs and adjustments for age and education where possible. This review sourced a wide range of studies including case-control studies. In addition, we report findings for differentiating dementias from NC and MCI from dementias, however, these were not considered to be an appropriate use of the MoCA.

The subject of the review assumes heterogeneity and therefore meta-analysis/es was not conducted. Quality ratings, forest plots of validated studies (sensitivity and specificity) with covariates (suggested cut-offs, age, education and country) and summary received operating characteristic curve (sROC) are presented.

The results showed a wide range in suggested cutoffs for MCI cross culturally, with variability in levels of sensitivity and specificity ranging from low to high. Poor methodological rigour appears to have affected reported accuracy and validity of the MoCA.

The review highlights the necessity for cross cultural considerations when using the MoCA, and recognizing it as a screen and not a diagnostic tool. Appropriate cutoffs and point adjustments for education are suggested.

Introduction

Mild cognitive impairment (MCI) is conceptualized as a transitional stage between normal aging and dementia and is recognized as a risk factor for AD [1], prompting many researchers to screen for MCI in order to provide early treatment and reduce the risk of progression to dementia [2]. Transitional markers from MCI to dementias are challenging to define and cognitive impairment may result from conditions not related to dementia (e.g. delirium, substance use, psychiatric illness and metabolic conditions). Factors including high premorbid intellectual function, high quality occupational and educational attainment; are suggested to provide a 'cognitive reserve' [3] conserving existing cognitive abilities even in the presence of biological risk. Due to the lack of stable biomarkers of neurodegeneration [4,5] diagnosis is reliant on clinical judgment, defined by clinical, cognitive, and functional criteria. Broadly, criteria for MCI relate to concern regarding a change in cognition, impairment in one or more cognitive domains and preservation of independence in functional abilities [6], however differences in clinical definitions exist [6,7] with clinical subtypes such as amnestic and non-amnestic MCI, and MCI due to deficits in single and multiple cognitive domains [8]. Indeed, MCI does not necessarily indicate a progressive degeneration with reversion rates from MCI to NC estimated at approximately 18% [9] and stability rates between 37% to 67% [10]; higher than rates of progression to dementia. Wide variation of incidence and prevalence rates exists for subtypes of MCI ([11,12], related to the differences in diagnostic criteria applied and how they are operationalized (e.g. cognitive assessments used, cut-offs for age and education) [11].

To facilitate the detection of MCI, many health professionals around the world are using the Montreal Cognitive Assessment (MoCA) as a brief cognitive screen across a variety of clinical settings. The MoCA, developed and validated by Nasreddine et al. [13], is a brief and potentially useful screening tool with high sensitivity and specificity for detecting MCI in persons performing in the normal range on the Mini-Mental State Examination (MMSE). It has been translated into 36 languages and additional alternative forms, yet only a small number have been validated and nearly all versions lack population-based data of individuals aged 80-plus [14]. For instance, the norms and cut-off scores for MCI and dementia of the MoCA are different among five Chinese versions [14] and the cut-off of 26 (+1 for <12 years education; recommended in the original study), has been suggested on many forms without validation.

Cross cultural cognitive assessment

Cultural factors (i.e. the accumulation of shared knowledge, reflected in behaviours, experience, beliefs, values, attitudes: [15]) and lifestyle factors (i.e. pattern of everyday life given available resources e.g. employment, diet, activities, living arrangements; [16]) will vary across and within groups. Education can be considered an aspect of culture where learning and schooling is embedded with the culture [17]. Culture may affect the validity of the cognitive tests used to identify MCI through biases related to the test or administration. For instance, where the construct an item is measuring is ambiguous due to poor translation or lack of cultural equivalent, or where there is unfamiliarity with both testing and stimuli or the related skill does not exist within a cultural repertoire [18].

Because of cultural background and lifestyle differences in Eastern and Western countries, it is necessary to assess the scale in patients from different cultures. The

MoCA is a screening instrument that evaluates seven cognitive domains on a single page and scores range from 0 to 30[13]. The domains are: visuospatial/executive functions, naming, verbal memory registration and learning, attention, abstraction, 5-minute delayed verbal memory, and orientation. Validation of the MoCA in different languages has led to linguistic and cultural modifications. For instance, in the Filipino version of the MoCA, an owl, the image that locals can recognize, is used as the replacement of rhinoceros [19]. The Chinese Beijing Version for example replaced Roman alphabets with Chinese character sequences in Alternating Trail Making and the names were changed to more common Chinese names to reflect local familiarity in the sentence repetition [20]. In the Korean version it is notable that the phonemic fluency task is replaced with a semantic fluency task [21]. Considering that frequency of phonemes may differ across languages, the scores of phonemic fluency task are not transferable from one country to another [22]. Indeed, the format for some cultures has been drawn into question, for instance trail making and cube copying test may be unfamiliar [23] and sentence repetition where revisions are not considered may not be equivalent becoming overly complex in translation. In contrast some items in the English version may not be sensitive enough, for instance Chinese speakers tend to perform better on digit span tasks [24].

Issues of normal variability in cognitive functioning. for instance, individuals at the polar ends of a normal distribution (e.g. 6% of healthy people will score at or below 1.5 standard deviations below the norm), may lead to a misclassification -false positive (FP) or false negative (FN)- based on a cut off. A normative approach was attempted by Wong et al [25] using the MOCA (Hong Kong version) and estimated high FP rates among NC highlighting the elevated risk of misclassification when using cut offs especially for those older and less educated.

The use of universal cut-offs also leads to a sampling bias (i.e. sample for which the cut off was established does not match with the individual being assessed) which, if uncontrolled, leads to false conclusions about an individual's cognitive ability. This may not relate to the cognitive test itself, but how a health care professional may interpret these results and how this may inform clinical decision making. A 'diagnostic cut-off' suggests that the cut-off alone can be used to interpret the MoCA rather than considering other factors affecting performance, or being aware that the MoCA is a screen and does not replace more thorough diagnostic assessment. However, where resources and training are limited, there may be a reliance on tools which are cost and time effective. Therefore, education and culturally adjusted 'cut-offs' can inform clinical decision making. Furthermore, different domains of cognition have cultural and language loadings. Language and literacy vary within and across cultures and will affect measurement of attention and memory. Performance on tests assessing executive functions are strongly related to education and culture [26,27] associated to processes of learning, language skills; perceptual skills and reasoning. As a consequence, individuals are unlikely to perform consistently across the MoCA due to factors other than cognitive decline. Even where the individual may speak English (and tested in English), English proficiency and language and cultural acculturation, may affect performance. In addition, adjustments to assessment, for instance using an interpreter introduces additional methodological issues.

Studies have demonstrated that the culturally and linguistically adapted versions of the MoCA are superior to the MMSE in screening for MCI [20,28,29], however, due to the differences in Western culture and difficulty in separating the influence of culture from education level, further revision of MoCA adaptations are required to tailor the

contents more accurately to local populations across the globe as well as optimizing cut-offs for detecting MCI.

Rationale

We aim to explore the cross cultural applicability of the measure and suggested cut-offs in more detail. Given that the MoCA is regularly used as a reference standard (i.e. selected as the measure against which other tests are evaluated or to inform classification), we aim to highlight considerations specific to language, cultural regions, age and education when using this measure clinically and as a research tool especially when assessing non western - English speaking individuals.

The MoCA was originally validated to identify MCI [13]. Due to the variation in classification there is no universally accepted "gold standard" criterion for diagnosis of MCI. As such this review considers the validations of the MoCA for MCI against recognised diagnostic criteria outlined in the inclusion criteria.

Data differentiating between MCI and dementia, and between dementia and NC are presented, however, these were considered to have limited utility. Cognitive screens alone should not be used in isolation to diagnose dementia. Guidelines [30,31]indicate the need for comprehensive physical, medical and cognitive assessment and history taking when it comes to diagnosing forms of dementia. Differentiating MCI and dementia is informed by identification of cognitive decline interfering with performing complex activities. The same applies to differentiating dementia from NC. None the less, dementia diagnosis based on MoCA has recently been reviewed [32], suggesting the quality of the studies was "not good enough" to conclusively recommend its use and that a lower cut-off was likely to be more accurate. While the review reported

optimal sensitivity and specificity due consideration was not given to education or variation in suggested cut-offs. Many studies have emerged since their search in 2012, and from a cross cultural perspective, given that these studies are validating the MoCA in different populations, these data have been included for reference.

Methods

The review protocol was registered with PROSPERO (registration no. CRD42016032837). Review followed PRISMA guidelines [33].

Search method for inclusion of studies

Published and unpublished studies were considered, no language or date restrictions were applied. Search terms were compiled into two concepts 1. (Montreal adj cognitive adj assessment) or MoCA 2. validat$ or norm$ or adapt$ or translat$ or cultur$ or reliab$).

Searches were conducted using the following databases: Ovid MEDLINE In-Process & Other Non-Indexed Citations and Ovid MEDLINE, Ovid PsycINFO and Ovid Embase (all years to 09 January 2016). Articles citing the original MoCA paper [13] were screened separately using Google Scholar.

Inclusion / Exclusion criteria

Criteria for inclusion of studies in the review were (1)dementia or MCI with comparison group, (2)more than 10 cases in each group, and (3)use of MoCA compared with a reference standard. The reference standard was defined as the Diagnostic and Statistical Manual of Mental Disorders (DSM) versions III to 5 [34], International Classification of Diseases – 10th edition (ICD-10) [35], the National Institute of

Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association Alzheimer's criteria (NINCDS-ADRDA) [36], Petersen's criteria [37], National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherché et l'Enseignement en Neurosciences (NINDS-AIREN) [38]or expert diagnosis following interview such as the Clinical Dementia Rating (CDR) [39].

Studies were exclude where MoCA score was used as a grouping variable, (2) where studies investigated English speaking, western samples and (3) where there was evidence of co-morbid mental health conditions e.g. schizophrenia, in the samples.

Selection of studies

Review authors independently reviewed titles and abstracts for inclusion criteria. Any questions regarding eligibility were resolved by seeking additional information from study authors and through discussion with the other author. Fig. 1. outlines a PRISMA flow diagram of the systematic review.

Insert Figure 1: PRISMA flow diagram of the systematic review.

Data extraction

From the selected papers, the following data were extracted: population, recruitment strategy, specification of illness, comparator (diagnostic standard), age, country, education adjustments, cut-off score, sensitivity and specificity (full details in supplementary data). Study authors were contacted for additional information where necessary. Reasonable attempts were made to translate papers to English (authors sought out translation by native speakers within their affiliated institutions).

Risk of bias across studies

Review authors independently assessed the methodological quality of the included studies using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) instrument [40], designed to evaluate the risk of bias and applicability of diagnostic accuracy studies. Any disagreements were decided by consensus.

The methodological quality in each of the domains (patient selection, execution of the index test and the reference standard, and flow of patients) was often difficult to assess as the required information was not clearly stated in the published reports. The QUADAS-2 scores for each domain and for each sub analysis are included (Supplementary Figure 1, 2 and 3), a summary is present in Figure 2.

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

For *Patient Selection*, case-control design studies were included for all analyses, where controls were thought to be a representative sample of the population which produced the cases. However, the case control design leads to spectrum bias (where sensitivity or specificity of the test differs with different patient characteristics or disease features e.g. severity) resulting in overestimation of test accuracy, which therefore resulted in a high risk of selection bias (differences between baseline

characteristics of the groups that are compared) and is recognised as a source of heterogeneity in this review.

In the *Index Test* domain over half of the studies were judged as being of unclear risk of bias, mainly due to the absence or unclear reporting on test administration, masking and pre-specification of thresholds. These studies were considered to have low risk of concerns about applicability.

In the *Reference Standard* domain all studies used a recognised reference standard that was likely to correctly classify the condition. However, the studies marked as unclear did not report the use of masking clearly and/or how the reference standard was operationalised and applied. Over 85% of the studies were assessed as being of low risk of concern about applicability.

In the *Flow and Timing* domain, less than a quarter of studies were judged as unclear, mainly due to absence of relevant information about the time intervals between index tests and reference standards and a lack of reporting about whether all participants were included in the final analyses.

Data Synthesis

Sensitivity, specificity for optimal cut offs (as indicated by the study authors) and sample size data were extracted into Review Manager 5.2 [41] for analysis. Where data were stratified to account for education or age we selected these data. Where exact sample sizes were not specified the whole group sample was divided for both groups (clinical and control). Analyses were separated into three categories:

differentiating MCI from NC, differentiating MCI from dementia, and differentiating dementia from NC. True (TP) and false (FP) positives, and true (TN) and false (FN) negatives were calculated for all studies. Sensitivity and specificity with 95% confidence intervals are presented separately for MCI and dementias.

Given the majority of studies were case control studies, accurate estimates of prevalence could not be calculated, limiting further analyses. To overcome the heterogeneity due to cut-offs, the accuracy data were pooled by fitting a summary received operating characteristic curve (sROC).

Results

Insert Figure 3: sROC curve for studies differentiating between MCI and NC.

Insert Figure 4: Forest plot for differentiating between MCI and NC.

MCI vs. NC

Twenty six studies in total (N=19060) investigated the ability of the MoCA to differentiate MCI from NC. Sensitivity ranged from 55% (at a specificity of 76%) to 96% (at a specificity of 76%). Specificity ranged from 19% (at sensitivity of 96%) to 97% (at sensitivity of 91%). A symmetrical sROC curve was generated with studies given equal weight for analysis (Figure 3). Studies, covariates and forest plots are presented in Figure 4.

While the number of studies and variation with the regions affect pooling of data, we divided up the studies into geographical regions to summarize the descriptive data. It does not suggest homogeneity between regions. Where indicated in the studies, education or age adjustments are reported.

China, Malaysia and Taiwan

Thirteen studies (N=16972) were validated across Chinese and Malay cohorts differentiating MCI from controls. These regions have been combined due to cross over in ethnicities involved in these studies. There were a number of validated adapted versions of the MoCA for this region: Beijing, Cantonese, Changsha, Hong Kong and Singapore. Suggested cut-offs ranged from 13 to 26 to differentiate MCI from NC. Two studies employing large samples, suggest age [14] and education [23] adjusted cut offs. Only one study investigated vascular cognitive impairment [42] suggesting a cutoff 26/27. Sensitivity ranged from 55% (at a specificity of 76%) to 96% (at a specificity of 76%). Specificity ranged from 62% (at sensitivity of 75%) to 97% (at sensitivity of 91%).

Lu et al. (2011) [23] noted significant effects of education, age, urban or rural residence, and sex on MoCA scores. Zhou et al. (2014) [43] raised concerns about applicability of MoCA-Cantonese in their low education sample (sensitivity: 49%). A number of studies retained the original suggestion of 1 point adjustment for <12 years education [14,23,43–47]. An education adjustment for <6years of 1 point was made by a number of studies [20,42,48,49] while Tan et al [14] suggested a 2 point adjustment for ≤6 years of education and Chu et al. [48] 2 points for illiteracy.

Two studies (N=489) specifically investigated the MoCA-Singapore [50,51] both adjusted cut-offs for education at different levels however accuracy levels were poor with high FP and FN. Sensitivity ranged from 54% (at a specificity of 51%) to 96% (at a specificity of 31%). Specificity ranged from 19% (at sensitivity of 96%) to 57% (at sensitivity of 59%).

The MoCA-Taiwan was only validated to differentiate between dementia and NC.

South East and Southern Asia (Japan, Korea, Philippines, Sri Lanka, Thailand)

One study (N=66) differentiating MCI from controls in Japan [52] suggested a cut off of 25/26, in an educated sample, no education adjustment was used. One study (N=85) in Thailand [53], suggested a cut off of 24/25 in a poorly educated sample. One study (N=152) in Korea [21] suggested a cutoff of 22/23 with a 1 point adjustment for <6 years of education. Julayanont et al [53] made an adjustment of 1 point for < 4 years education and an additional 1 point for meeting the illiteracy criteria. All studies displayed high sensitivity and specificity, but these results were based on small samples.

The MoCA- Philippines and the MoCA- Sri Lanka were only validated to differentiate between dementia and NC.

Europe (Germany, Holland, Poland, Portugal)

One study from each region (N=639) validated the MoCA differentiating MCI from controls. Where education level was broadly comparable the suggested cutoffs were 25/26, in the lower education Portuguese sample [28] the suggested cut off was 22.

Sensitivity ranged from 72% (at a specificity of 73%) to 93% (at a specificity of 62%). Specificity ranged from 54% (at sensitivity of 81%) to 77% (at sensitivity of 81%).

While the Polish [54] and German [55] studies retained the original 1 point adjustment, the Dutch study [56] implemented a 2 point adjustment for ≤ 12 years of education. Freitas et al. [23] did not include an adjustment for their sample.

Additionally, high FP rates were noted in the German, Polish and Dutch studies with lower end confidence intervals for specificity approaching or falling below chance.

Middle East (Israel and Turkey)

One study from each region validated the MoCA differentiating MCI from controls.

Kaya et al [29] (N=196) reported education adjusted cutoffs for the MoCA-Turkey, however there were high FP rates with specificity around chance level, sensitivity was also low for those less educated.

The MoCA-Hebrew study (N=154) [57] suggested a cut-off of 26. Education adjustments were not used in these studies.

South America (Brazil, Chile, Columbia)

Two studies (N=168) validated the MoCA differentiating MCI from controls in Brazil [58,59] both reporting a cut off of 25 with adequate sensitivity and specificity. One study [60] (N=68) validated the MoCA-Chile suggesting a cut off of 24 in an educated sample. However the FP rate was high, with wide confidence intervals. In the Columbian sample [61] (N=216) suggested a cutoff of 23 in an educated sample with adequate sensitivity and specificity.

Memoria et al. [59] retained the original 1 point adjustment, while Gil et al [61] suggested it should be for < 11 years education.

Insert Figure 5: Forest plot for studies differentiating between MCI and dementia.

Insert Figure 6: Forest plot for studies differentiating between dementia and NC.

Dementia vs. NC

Eighteen studies in total (N=12670) investigated the ability of the MoCA to differentiate dementia from NC. Studies, covariates and forest plots are presented in Figure 5. Dementia types: Alzheimer's disease (10), Dementia with Lewy bodies (1), mixed type dementia (4), behavioral variant fronto-temporal dementia (1), MCI and dementia - MCI+ (1), vascular dementia (1), major NCD (1). Cutoffs ranged from 13 to 26. Sensitivity ranged from 75% (at a specificity of 53%) to 100% (at a specificity of 89%). Specificity ranged from 57% (at sensitivity of 75%) to 100% (at sensitivity of 94%).

MCI vs. Dementia

Seven studies in total (N=1087) investigated the ability of the MoCA to differentiate dementia from NC. Studies, covariates and forest plots are presented in Figure 6. Cutoffs ranged from 15 to 25. Sensitivity ranged from 81% (at a specificity of 80%) to 100% (at a specificity of 96%). Specificity ranged from 70% (at sensitivity of 89%) to 98% (at sensitivity of 96%).

Heterogeneity

The studies included in the review are heterogeneous, with different demographic samples, recruitment settings, education levels, and variations in study population

(e.g. severity of disease and co-morbidities). The study specific cut-offs adjust for the above factors within their sample populations. However, there are likely other factors such as age, education, spectrum (severity or stage) of the target disorder, operator characteristics (for index test and reference standard) and technical features (including different versions of the test and item adjustment) that have not been accounted for. Diagnostic criteria used will lead to diagnostic variability across studies in identifying MCI and subsequently affect the validity and reliability, sensitivity and specificity of the MOCA in identifying MCI.

We assessed potential heterogeneity through visual inspection of risk of bias, descriptive data and forest plots. We specified two factors that may contribute to heterogeneity explored by relating study level co-variates: quality item - selection of patients, and education. These were assessed effect by plotting sROCs by covariate. We compared high, unclear and low risk studies for the item, selection of patients, in sROC (summary of pairs of sensitivity & specificity in ROC space). These data suggest that accuracy varies with selection of studies with high/unclear/low (Figure 7). Studies at high risk of bias reported higher level of accuracy than low risk studies and unclear studies reported lower accuracy.

Due to different reporting of education attainment, heterogeneity was assessed through visual inspection of forest plot (low to high education). It appeared that there was a mild distinction between primary and secondary educated (at the 8 year point). Below 8 years there was no clear pattern of improving sensitivity/specificity as hypothesized (pooled sensitivity/specificity: 74%/70%). For >8 years sensitivity appears stable while specificity fluctuates (pooled sensitivity/specificity: 86%/70%). As such there was no clear pattern for the effect of education.

Insert Figure 7: sROC curve displaying influence of quality item, selection of patients on accuracy.

Discussion

The investigation of the cross cultural applicability of the MoCA in screening for MCI, revealed wide variation in cut-offs both across countries and within region and as a result it is not possible to conclusively suggest adjusted cut-offs. Apart from China, there were a low number of studies, generally only one, validating the measure in each region. Many translated forms of the MoCA have not been validated, and there are no published studies validating its use in large populations e.g. use in different countries in Africa or the Middle East. A number of studies were not considered to display adequate sensitivity with high FP rates [29,50,51,54]. However, many of the studies indicated adequate to high sensitivity and specificity against study specific cut-offs and education adjustments.

No clear pattern was identified for the effect of education. Years of education may not be equivalent across countries, where quality of education varies. Formal educational levels will differ between countries, where variation will exist in the organization of educational systems, access, learning outcomes and resources (e.g. teachers). Indeed, literacy in later life (as a measure of educational quality) is a stronger predictor of cognitive performance in older individuals than years of education [62]. In addition, certain items on the MoCA are likely to be influenced by education [43], for instance, clock drawing, phonemic verbal fluency, and verbal abstraction [59].

The reviewed studies were primarily case control studies, with small sample sizes, with some validations only carried out for dementia. As expected, the case-control

studies (high risk of bias) appeared to over-estimate sensitivity and specificity, with studies with unclear risk of bias for selection of patients reporting lower accuracy (than low risk studies), possibly indicative of poor methodological quality. Item adjustments may also have effected scores with different forms being used across studies, this is especially pertinent to China where multiple versions exist. In this review, variation in both the diagnostic criteria and operational definitions precluded us from making quantitative interpretations of the influence of both factors. However given that these variables substantially affect estimation of prevalence rates [11] it is likely that they will exert a strong influence on a comparison between studies.

A comprehensive review of the literature was conducted with attempts made to obtain translations of non-English papers and contact with authors was made where further information or additional data were required. Samples from studies conducted in the context of different specialist settings were included in this review as they were considered to have ecological validity applicable to clinicians in care settings, however, it is recognized that this gives rise to heterogeneity. The diversity of countries that have studied the MoCA both between and within cultures adds to the generalizability of the results. The review does not cover the influence of acculturation and use of interpreters which are likely to affect administration and reliability of scores. In a study of acculturation effects in North America, high FP rates and lower cut-offs were reported (even where English language was proficient) [63]. When using MoCA applicable to their country of origin, it's necessary to consider acculturation (positive and negative effects), regional variation e.g. urban/rural, regional language variation and quality of education.

Given the aging immigrant population across western countries and the anticipated increase in dementia prevalence worldwide in coming decades, earlier and more accurate detection of MCI in these populations is important. Although the MoCA is an effective brief cognitive assessment tool that has shown strong psychometric properties in several countries and shows particular advantage in culturally and linguistically diverse populations, our results highlight the need for caution when using MoCA as a screen for MCI without appropriately accounting for the effects of education, language and cultural diversity. Hu et al. [20] found several cultural differences in practice in a Chinese version of MoCA. Subjects with lower levels of education did not understand how to perform the 'Alternating Trail Making' test. This has also been reported in a study in subjects from Hong Kong [64]. In the 'Naming tests', 58.2% of subjects were unable to name the 'rhinoceros' and 'camel' since the animal is not within the realm of their general understanding or vocabulary. In the word memory test, the original words are not easy to memorize for Chinese people especially the words 'velvet' and 'church', which fall outside Chinese cultural background. In the 'Abstract thinking' test the 'watch' and 'ruler' were only correctly remembered by 41.8% of subjects, and even among those of a higher level of education they were only correctly identified 50% of the time [20].

Due to the differences in cultures, further revision is required to tailor the contents more accurately to local populations and to highlight the difficulty in separating the influence of culture from education level. In immigrant populations additional caution should be given to interpretation of scores when using forms for the country of origin or English language forms due to acculturation effects.

It is also necessary to redefine cut-off rates based on age and education-specific normative values so that individual test items can be modified appropriately, enabling

better content validity (i.e. type and quality of the test items to assess mild impairments in cognitive functioning). There is need for future research to provide consistent data reporting conventions to facilitate direct comparison across countries. Future research, may also make use of education classification systems (e.g. International Standard Classification of Education [65] to facilitate more accurate cross cultural comparisons of the effects of education. Validations need to be carried out on different versions, especially those available on the MoCA website as these can appear to be endorsed by the creators. Further research could also explore the effect of acculturation on MoCA performance as this is especially pertinent given the aging migratory populations. Identification of salient domains which show cultural variability or have greater/less reliability could also inform development of MoCA forms for unvalidated populations.

Conclusions

The findings of this review suggest that it would be more appropriate to use cutoffs and point adjustments for education as suggested by validation authors (not those of the original paper). The 'low' quality of validation papers and likelihood that accuracy is over estimated highlighted the need for clinicians and research to use the MoCA appropriately, that is, as a brief, cognitive screening tool. Furthermore, it is important to stress that test scores should be interpreted in light of other clinical data, such as clinical history, collateral information, and behavioral observations [25]. It should be noted that although the MoCA has been adapted across languages and cultures, it takes little account of acculturation factors. Utilizing cognitive screens can be implemented to obtain objective data, however, the review indicates the limitations and

suitability of using versions of the MOCA within the context of individual and cultural factors.

In addition, the reported range of cut-offs in validated papers suggests that there is questionable utility in using unvalidated measures and direct translations.

References

[1]   Levey A, Lah J, Goldstein F, Steenland K, Bliwise D (2006) Mild cognitive impairment: An opportunity to identify patients at high risk for progression to Alzheimer's disease. *Clin. Ther.* **28**, 991–1001.

[2]   Ritchie K, Artero S, Touchon J (2001) Classification criteria for mild cognitive impairment: a population-based validation study. *Neurology* **56**, 37–42.

[3]   Stern Y (2012) Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol.* **11**, 1006–1012.

[4]   Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FRJ, Visser PJ, Aalten P, Aarsland D, Alcolea D, Alexander M, Almdahl IS, Arnold SE, Baldeiras I, Barthel H, van Berckel BNM, Bibeau K, Blennow K, Brooks DJ, van Buchem MA, Camus V, Cavedo E, Chen K, Chetelat G, Cohen AD, Drzezga A, Engelborghs S, Fagan AM, Fladby T, Fleisher AS, van der Flier WM, Ford L, Förster S, Fortea J, Foskett N, Frederiksen KS, Freund-Levi Y, Frisoni GB, Froelich L, Gabryelewicz T, Gill KD, Gkatzima O, Gómez-Tortosa E, Gordon MF, Grimmer T, Hampel H, Hausner L, Hellwig S, Herukka S-K, Hildebrandt H, Ishihara L, Ivanoiu A, Jagust WJ, Johannsen P, Kandimalla R, Kapaki E, Klimkowicz-Mrowiec A, Klunk WE, Köhler S, Koglin N, Kornhuber J, Kramberger MG, Van Laere K, Landau SM, Lee DY, de Leon M, Lisetti V, Lleó A, Madsen K, Maier W, Marcusson J, Mattsson N, de Mendonça A, Meulenbroek O, Meyer PT, Mintun MA, Mok V, Molinuevo JL, Møllergård HM, Morris JC, Mroczko B, Van der Mussele S, Na DL, Newberg A, Nordberg A, Nordlund A, Novak GP, Paraskevas GP, Parnetti L, Perera G, Peters O, Popp J, Prabhakar S, Rabinovici GD, Ramakers IHGB, Rami L, Resende de Oliveira C, Rinne JO, Rodrigue KM, Rodríguez-Rodríguez E, Roe CM, Rot U, Rowe CC, Rüther E, Sabri O, Sanchez-Juan P, Santana I, Sarazin M, Schröder J, Schütte C, Seo SW, Soetewey F, Soininen H, Spiru L, Struyfs H, Teunissen CE, Tsolaki M, Vandenberghe R, Verbeek MM, Villemagne VL, Vos SJB, van Waalwijk van Doorn LJC, Waldemar G, Wallin A, Wallin ÅK, Wiltfang J, Wolk DA, Zboch M, Zetterberg H (2015) Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia. *JAMA* **313**, 1924.

[5]   Saleem M, Herrmann N, Swardfager W, Eisen R, Lanctôt KL (2015) Inflammatory Markers in Mild Cognitive Impairment: A Meta-Analysis. *J. Alzheimer's Dis.* **47**, 669–679.

[6]   Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst

A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **7**, 270–279.

[7]   Petersen RC (2011) Mild cognitive impairment. *N. Engl. J. Med.* **364**, 2227–2234.

[8]   Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, Delis DC (2009) Quantification of Five Neuropsychological Approaches to Defining Mild Cognitive Impairment. *Am. J. Geriatr. Psychiatry* **17**, 368–375.

[9]   Canevelli M, Grande G, Lacorte E, Quarchioni E, Cesari M, Mariani C, Bruno G, Vanacore N (2016) Spontaneous Reversion of Mild Cognitive Impairment to Normal Cognition: A Systematic Review of Literature and Meta-Analysis. *J. Am. Med. Dir. Assoc.* **17**, 943–948.

[10] Pandya SY, Clem MA, Silva LM, Woon FL (2016) Does mild cognitive impairment always lead to dementia? A review. *J. Neurol. Sci.* **369**, 57–62.

[11] Ward A, Arrighi HM, Michels S, Cedarbaum JM (2012) Mild cognitive impairment: Disparity of incidence and prevalence estimates. *Alzheimer's Dement.* **8**, 14–21.

[12] Roberts R, Knopman DS (2013) Classification and Epidemiology of MCI. *Clin. Geriatr. Med.* **29**, 753–772.

[13] Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005) The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699.

[14] Tan J, Li N, Gao J, Wang L, Zhao Y, Yu B, Du W, Zhang W, Cui L, Wang Q, Li J, Yang J, Yu J, Xia X, Zhou P (2015) Optimal cutoff scores for dementia and mild cognitive impairment of the Montreal Cognitive Assessment among elderly and oldest-old Chinese population. *J. Alzheimers. Dis.* **43**, 1403–1412.

[15] Porter, R. E., Samovar LA (1994) An introduction to intercultural communication. In *Intercultural communication: A reader*, L.A. Samovar & R.E. Porter, ed. Wadsworth, Belmont, CA, pp. 4–26.

[16] Zapf W (1987) *Individualisierung und Sicherheit: Untersuchungen zur Lebensqualität in der Bundesrepublik Deutschland.*, Beck.

[17] Ardila A, Ostrosky-Solis F, Mendoza VU (2000) Learning to read is much more than learning to read: a neuropsychologically based reading program. *J. Int. Neuropsychol. Soc.* **6**, 789–801.

[18] van de Vijver F, Tanzer NK (2004) Bias and equivalence in cross-cultural assessment: an overview. *Rev. Eur. Psychol. Appliquée/European Rev. Appl. Psychol.* **54**, 119–135.

[19] Dominguez JC, Orquiza MGS, Soriano JR, Magpantay CD, Esteban RC, Corrales ML, Ampil ER (2013) Adaptation of the Montreal Cognitive Assessment for elderly Filipino patients. *East Asian Arch. Psychiatry* **23**, 80–85.

[20] Hu JB, Zhou WH, Hu SH, Huang ML, Wei N, Qi HL, Huang JW, Xu Y (2013) Cross-cultural difference and validation of the Chinese version of Montreal Cognitive Assessment in older adults residing in Eastern China: preliminary findings. *Arch. Gerontol. Geriatr.* **56**, 38–43.

[21] Lee J-YJ, Lee DW, Cho S, Na DL, Jeon HJ, Kim S, Lee YR, Youn J, Kwon M, Lee J-YJ, Cho MJ (2008) Brief screening for mild cognitive impairment in

elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. *J. Geriatr. Psychiatry Neurol.* **21**, 104–110.

[22] Tombaugh TN, Kozak J, Rees L (1999) Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Arch. Clin. Neuropsychol.* **14**, 167–77.

[23] Lu J, Li D, Li F, Zhou A, Wang F, Zuo X, Jia X-F, Song H, Jia J (2011) Montreal Cognitive Assessment in Detecting Cognitive Impairment in Chinese Elderly Individuals: A Population-Based Study. *J. Geriatr. Psychiatry Neurol.* **24**, 184–190.

[24] Hedden T, Park DC, Nisbett R, Ji L-J, Jing Q, Jiao S (2002) Cultural variation in verbal versus spatial neuropsychological function across the life span. *Neuropsychology* **16**, 65–73.

[25] Wong A, Law L, Liu W, Wang Z, Lo E, Lau A, Wong L, Mok V (2015) Montreal Cognitive Assessment: One Cutoff Never Fits All. *Stroke* **46**, 3547–3550.

[26] Friedman NP, Miyake A, Corley RP, Young SE, DeFries JC, Hewitt JK (2006) Not All Executive Functions Are Related to Intelligence. *Psychol. Sci.* **17**, 172–179.

[27] Lewis C, Koyasu M, Oh S, Ogawa A, Short B, Huang Z (2009) Culture, executive function, and social understanding. *New Dir. Child Adolesc. Dev.* **2009**, 69–85.

[28] Freitas S, Simões MR, Alves L, Santana I (2013) Montreal Cognitive Assessment. *Alzheimer Dis. Assoc. Disord.* **27**, 37–43.

[29] Kaya Y, Aki OE, Can UA, Derle E, Kibaro lu S, Barak A (2014) Validation of Montreal Cognitive Assessment and Discriminant Power of Montreal Cognitive Assessment Subtests in Patients With Mild Cognitive Impairment and Alzheimer Dementia in Turkish Population. *J. Geriatr. Psychiatry Neurol.* **27**, 103–109.

[30] Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, Thies B, Phelps CH (2011) Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **7**, 257–262.

[31] Pathway-Dementia NICE (2006) *Dementia: supporting people with dementia and their carers in health and social care NICE guidelines [CG42] Published date: November 2006 Last updated: April 2014.*

[32] Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S (2015) Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane database Syst. Rev.* **10**, CD010775.

[33] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **6**, e1000097.

[34] American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders*, Washington, DC.

[35] World Health Organization (1993) *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*.

[36] Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, Dickson DW, Duyckaerts C, Frosch MP, Masliah E, Mirra SS, Nelson PT, Schneider JA, Thal DR, Thies B, Trojanowski JQ, Vinters H V., Montine TJ (2012) National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's Dement.* **8**, 1–13.

[37]   Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* **256**, 183–94.

[38]   Roman GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, Amaducci L, Orgogozo J-M, Brun A, Hofman A, Moody DM, O'Brien MD, Yamaguchi T, Grafman J, Drayer BP, Bennett DA, Fisher M, Ogata J, Kokmen E, Bermejo F, Wolf PA, Gorelick PB, Bick KL, Pajeau AK, Bell MA, DeCarli C, Culebras A, Korczyn AD, Bogousslavsky J, Hartmann A, Scheinberg P (1993) Vascular dementia: Diagnostic criteria for research studies: Report of the NINDS-AIREN International Workshop. *Neurology* **43**, 250–250.

[39]   Morris JC (1993) The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* **43**, 2412–2412.

[40]   Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, QUADAS-2 Group (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **155**, 529–36.

[41]   (2012) Review Manager (RevMan).

[42]   Tu Q-Y, Jin H, Ding B-R, Yang X, Lei Z-H, Bai S, Zhang Y-D, Tang X-Q (2013) Reliability, validity, and optimal cutoff score of the montreal cognitive assessment (changsha version) in ischemic cerebrovascular disease patients of hunan province, china. *Dement. Geriatr. Cogn. Dis. Extra* **3**, 25–36.

[43]   Zhou S, Zhu J, Zhang N, Wang B, Li T, Lv X, Ng TP, Yu X, Wang H (2014) The Influence of Education on Chinese Version of Montreal Cognitive Assessment in Detecting Amnesic Mild Cognitive Impairment among Older People in a Beijing Rural Community. *Sci. World J.* **2014**, 1–7.

[44]   Dong Y, Lee WY, Hilal S, Saini M, Wong TY, Chen CL-H, Venketasubramanian N, Ikram MK (2013) Comparison of the Montreal Cognitive Assessment and the Mini-Mental State Examination in detecting multi-domain mild cognitive impairment in a Chinese sub-sample drawn from a population-based study. *Int. Psychogeriatrics* **25**, 1831–1838.

[45]   Ng KHY (2008) The validity of the Montreal cognitive assessment (Cantonese version) as a screening tool for mild cognitive impairment in Hong Kong Chinese.

[46]   Yu J, Li J, Huang X (2012) The Beijing version of the Montreal Cognitive Assessment as a brief screening tool for mild cognitive impairment: a community-based study. *BMC Psychiatry* **12**, 156.

[47]   Zhao S, Liu G, Shen Y, Zhao Y (2011) Reasonable neuropsychological battery to identify mild cognitive impairment. *Med. Hypotheses* **76**, 50–53.

[48]   Chu L, Ng KHY, Law ACK, Lee AM, Kwan F (2015) Validity of the Cantonese Chinese Montreal Cognitive Assessment in Southern Chinese. *Geriatr. Gerontol. Int.* **15**, 96–103.

[49]   Yeung PY, Wong LL, Chan CC, Leung JLM, Yung CY (2014) A validation study of the Hong Kong version of Montreal Cognitive Assessment (HK-MoCA) in Chinese older adults in Hong Kong. *Hong Kong Med. J.* **20**, 504–510.

[50]   Ng A, Chew I, Narasimhalu K, Kandiah N (2013) Effectiveness of Montreal Cognitive Assessment for the diagnosis of mild cognitive impairment and mild Alzheimer's disease in Singapore. *Singapore Med. J.* **54**, 616–619.

[51]   Ng TP, Feng L, Lim WS, Chong MS, Lee TS, Yap KB, Tsoi T, Liew TM, Gao Q, Collinson S, Kandiah N, Yap P (2015) Montreal Cognitive Assessment for Screening Mild Cognitive Impairment: Variations in Test Performance and Scores by Education in Singapore. *Dement. Geriatr. Cogn. Disord.* **39**, 176–

185.

[52]   Fujiwara Y, Suzuki H, Yasunaga M, Sugiyama M, Ijuin M, Sakuma N, Inagaki H, Iwasa H, Ura C, Yatomi N, Ishii K, Tokumaru AM, 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. *Geriatr. Gerontol. Int.* **10**, 225–232.

[53]   Julayanont P, Tangwongchai S, Hemrungrojn S, Tunvirachaisakul C, Phanthumchinda K, Hongsawat J, Suwichanarakul P, Thanasirorat S, Nasreddine ZS (2015) The Montreal Cognitive Assessment-Basic: A Screening Tool for Mild Cognitive Impairment in Illiterate and Low-Educated Elderly Adults. *J. Am. Geriatr. Soc.* **63**, 2550–2554.

[54]   Magierska J, Magierski R, Fendler W, Kloszewska I, Sobow TM (2012) Clinical application of the Polish adaptation of the Montreal Cognitive Assessment (MoCA) test in screening for cognitive impairment. *Neurol. Neurochir. Pol.* **46**, 130–139.

[55]   Costa AS, Fimm B, Friesen P, Soundjock H, Rottschy C, Gross T, Eitner F, Reich A, Schulz JB, Nasreddine ZS, Reetz K (2012) Alternate-form reliability of the Montreal cognitive assessment screening test in a clinical setting. *Dement. Geriatr. Cogn. Disord.* **33**, 379–384.

[56]   Thissen AJAM, van Bergen F, de Jonghe JFM, Kessels RPC, Dautzenberg PLJ (2010) [Applicability and validity of the Dutch version of the Montreal Cognitive Assessment (moCA-d) in diagnosing MCI]. *Tijdschr. Gerontol. Geriatr.* **41**, 231–40.

[57]   Lifshitz M, Dwolatzky T, Press Y (2012) Validation of the Hebrew version of the MoCA test as a screening instrument for the early detection of mild cognitive impairment in elderly individuals. *J. Geriatr. Psychiatry Neurol.* **25**, 155–161.

[58]   Martinelli JE, Cecato JF, Bartholomeu D, Montiel JM (2014) Comparison of the diagnostic accuracy of neuropsychological tests in differentiating Alzheimer's disease from mild cognitive impairment: can the montreal cognitive assessment be better than the cambridge cognitive examination?. *Dement. Geriatr. Cogn. Dis. Extra* **4**, 113–121.

[59]   Memoria CM, Yassuda MS, Nakano EY, Forlenza O V (2013) Brief screening for mild cognitive impairment: validation of the Brazilian version of the Montreal cognitive assessment. *Int. J. Geriatr. Psychiatry* **28**, 34–40.

[60]   Munoz-Neira C, Henriquez Chaparro F, Delgado C, Brown J, Slachevsky A (2014) Test Your Memory-Spanish version (TYM-S): a validation study of a self-administered cognitive screening test. *Int. J. Geriatr. Psychiatry* **29**, 730–740.

[61]   Gil L, Ruiz de Sánchez C, Gil F, Romero SJ, Pretelt Burgos F (2015) Validation of the Montreal Cognitive Assessment (MoCA) in Spanish as a screening tool for mild cognitive impairment and mild dementia in patients over 65 years old in Bogotá, Colombia. *Int. J. Geriatr. Psychiatry* **30**, 655–62.

[62]   Sisco S, Gross AL, Shih R a, Sachs BC, Glymour MM, Bangen KJ, Benitez A, Skinner J, Schneider BC, Manly JJ (2014) The Role of Early-Life Educational Quality and Literacy in Explaining Racial Disparities in Cognition in Late Life. *J. Gerontol. B. Psychol. Sci. Soc. Sci.* 1–11.

[63]   Goldstein FC, Ashley A V., Miller E, Alexeeva O, Zanders L, King V (2014) Validity of the Montreal Cognitive Assessment as a Screen for Mild Cognitive Impairment and Dementia in African Americans. *J. Geriatr. Psychiatry Neurol.* **27**, 199–203.

[64] Wong A, Xiong YY, Kwan PWL, Chan AYY, Lam WWM, Wang K, Chu WCW, Nyenhuis DL, Nasreddine Z, Wong LKS, Mok VCT (2009) The validity, reliability and clinical utility of the Hong Kong Montreal Cognitive Assessment (HK-MoCA) in patients with cerebral small vessel disease. *Dement. Geriatr. Cogn. Disord.* **28**, 81–87.

[65] UNESCO Institute for Statistics (UIS). (2012). *International Standard Classification of Education: ISCED 2011*. UIS, Montreal, Quebec.

Figure 1: PRISMA flow diagram of the systematic review.

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

Figure 3: sROC curve for studies differentiating between MCI and NC.

Study	TP	FP	FN	TN	Cut off value	Education	Country	Age	Sensitivity (95% CI)	Specificity (95% CI)
Martinelli 2014	37	3	8	36	25	6	Brazil	76.6 (7.06)	0.82 [0.68, 0.92]	0.92 [0.79, 0.98]
Memoria 2013	39	9	4	32	25	11.41 (4.23)	Brazil	65 to 90	0.91 [0.78, 0.97]	0.78 [0.62, 0.89]
Munoz-Neira 2014	13	9	1	21	24	12.79 (4.59)	Chile	71.93 (7.06)	0.93 [0.66, 1.00]	0.70 [0.51, 0.85]
Chu 2015	68	31	19	84	22/23	4.62 (5.19)	China	77.2 (6.3)	0.78 [0.68, 0.86]	0.73 [0.64, 0.81]
Dong 2013	66	10	17	118	19/20	3.0 (3.8)	China	74.3 (5.5)	0.80 [0.69, 0.88]	0.92 [0.86, 0.96]
Hu 2013	76	19	8	127	26	9.82 (3.0)	China	60.7 (5.0)	0.90 [0.82, 0.96]	0.87 [0.80, 0.92]
Liew 2015	17	8	4	27	23/24	≤ primary	China	71.8 (6.71)	0.81 [0.58, 0.95]	0.77 [0.60, 0.90]
Liew 2015a	11	8	9	26	24/25	≥ secondary	China	71.8 (6.71)	0.55 [0.32, 0.77]	0.76 [0.59, 0.89]
Lu 2011	1358	1100	329	5183	13/14; 19/20; 24/25	0; 1 to 6; ≥7	China	75.1 (0.9)	0.80 [0.79, 0.82]	0.82 [0.82, 0.83]
Ng 2008	43	19	11	55	23	3.91	China	78.74	0.80 [0.66, 0.89]	0.74 [0.63, 0.84]
Tan 2015	584	180	97	1050	25	7.5 (5.67)	China	60-79	0.86 [0.83, 0.88]	0.85 [0.83, 0.87]
Tan 2015a	1245	98	215	2551	24	7.5 (5.67)	China	80-89	0.85 [0.83, 0.87]	0.96 [0.96, 0.97]
Tan 2015b	50	3	5	97	23	7.5 (5.67)	China	≥90	0.91 [0.80, 0.97]	0.97 [0.91, 0.99]
Tsai 2012	65	8	6	30	23/24	8.9 (5.1)	China	79.2 (6.8)	0.92 [0.83, 0.97]	0.79 [0.63, 0.90]
Tu 2013	107	32	4	100	26/27	8.14 (4.24)	China	69.41 (11.22)	0.96 [0.91, 0.99]	0.76 [0.68, 0.83]
Yeung 2014	77	13	16	36	21/22	4.80 (4.78)	China	76.45 (7.54)	0.83 [0.74, 0.90]	0.73 [0.59, 0.85]
Yu 2012	79	312	36	553	22	8.4 (5.46)	China	71.45 (7.26)	0.69 [0.59, 0.77]	0.64 [0.61, 0.67]
Zhao 2011	116	15	34	135	23/24	na	China	70.67 (4.27)	0.77 [0.70, 0.84]	0.90 [0.84, 0.94]
Zhou 2014	18	56	6	92	20/21	6.17 (1.24)	China	67.17 (6.59)	0.75 [0.53, 0.90]	0.62 [0.54, 0.70]
Gil 2015	97	17	12	67	23	13 (5.1)	Columbia	65 (13.4)	0.89 [0.82, 0.94]	0.80 [0.70, 0.88]
Costa 2012	28	38	2	62	26	11.47 (2.85)	Germany	67.80 (8.13)	0.93 [0.78, 0.99]	0.62 [0.52, 0.72]
Thissen 2010	23	8	9	22	25	na	Holland	75.9 (6.4)	0.72 [0.53, 0.86]	0.73 [0.54, 0.88]
Lifshitz 2012	70	19	4	61	26	na	Israel	76.3 (5.61)	0.95 [0.87, 0.99]	0.76 [0.65, 0.85]
Fujiwara 2010	28	4	2	32	25/26	11.5 (3.1)	Japan	77.3 (6.3)	0.93 [0.78, 0.99]	0.89 [0.74, 0.97]
Lee 2008	33	18	4	97	22/23	8.3 (3.8)	Korea	70.4 (8.6)	0.89 [0.75, 0.97]	0.84 [0.76, 0.90]
Magierska 2012	34	17	8	20	25	13.4	Poland	74.2 (6.4)	0.81 [0.66, 0.91]	0.54 [0.37, 0.71]
Freitas 2013	73	41	17	139	22	6.50 (4.57)	Portugal	70.52 (7.95)	0.81 [0.71, 0.89]	0.77 [0.70, 0.83]
Ng 2013	24	42	1	10	27	>10	Singapore	62.43 (9.40)	0.96 [0.80, 1.00]	0.19 [0.10, 0.33]
Ng 2013a	23	36	1	16	26	≤10	Singapore	62.43 (9.40)	0.96 [0.79, 1.00]	0.31 [0.19, 0.45]
Ng 2015	17	25	10	30	21/22	0	Singapore	na	0.63 [0.42, 0.81]	0.55 [0.41, 0.68]
Ng 2015a	20	37	14	49	22/23	1 to 6	Singapore	na	0.59 [0.41, 0.75]	0.57 [0.46, 0.68]
Ng 2015b	25	43	21	45	27/28	>6	Singapore	na	0.54 [0.39, 0.69]	0.51 [0.40, 0.62]
Julayanont 2015	36	6	6	37	24/25	2.9 (1.7)	Thailand	70.2 (6.6)	0.86 [0.71, 0.95]	0.86 [0.72, 0.95]
Kaya 2014	25	14	13	13	17/18	5 to 8	Turkey	74.2 (8.83)	0.66 [0.49, 0.80]	0.48 [0.29, 0.68]
Kaya 2014a	28	12	10	15	20/21	9 to 11	Turkey	74.2 (8.83)	0.74 [0.57, 0.87]	0.56 [0.35, 0.75]
Kaya 2014b	31	11	7	17	22/23	≥12	Turkey	74.2 (8.83)	0.82 [0.66, 0.92]	0.61 [0.41, 0.78]

Figure 4: Forest plot for differentiating between MCI and NC. a, b indicates education or age adjustment within the same study.

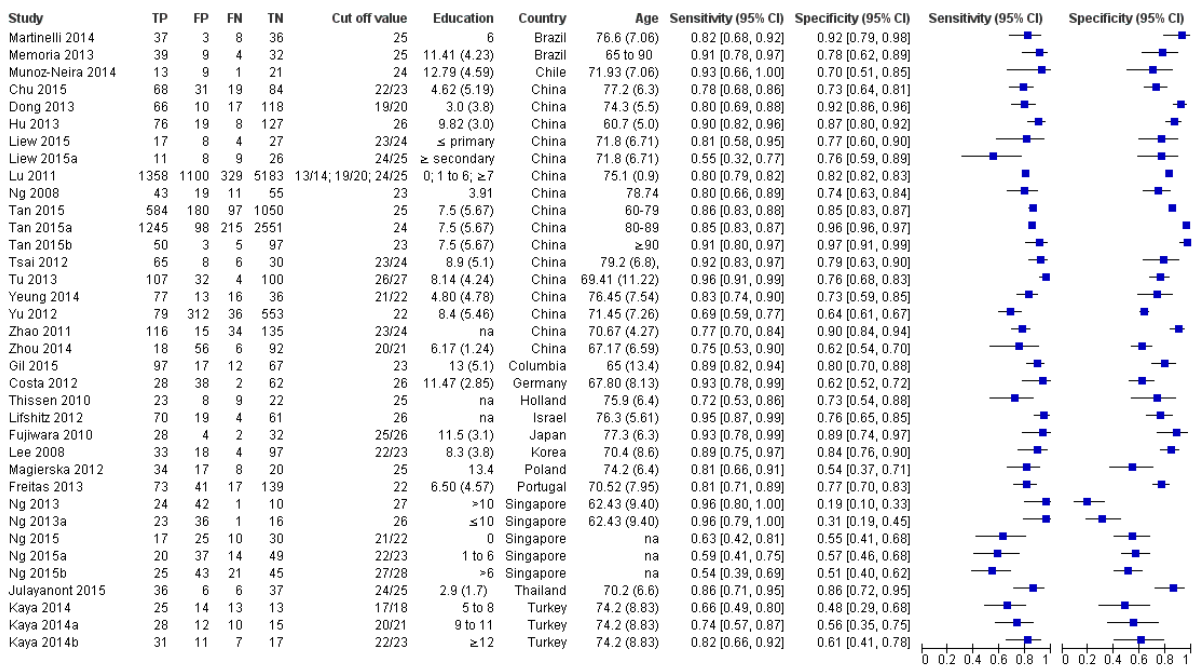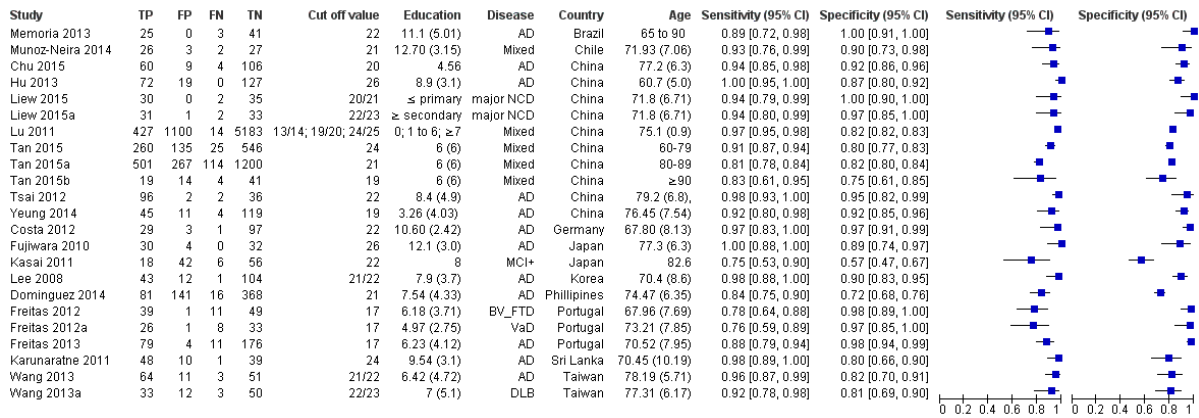| Study	TP	FP	FN	TN	Cut off value	Education	Disease	Country	Age	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Memoria 2013	25	0	3	41	22	11.1 (5.01)	AD	Brazil	65 to 90	0.89 [0.72, 0.98]	1.00 [0.91, 1.00]		
Munoz-Neira 2014	26	3	2	27	21	12.70 (3.15)	Mixed	Chile	71.93 (7.06)	0.93 [0.76, 0.99]	0.90 [0.73, 0.98]		
Chu 2015	60	9	4	106	20	4.56	AD	China	77.2 (6.3)	0.94 [0.85, 0.98]	0.92 [0.86, 0.96]		
Hu 2013	72	19	0	127	26	8.9 (3.1)	AD	China	60.7 (5.0)	1.00 [0.95, 1.00]	0.87 [0.80, 0.92]		
Liew 2015	30	0	2	35	20/21	≤ primary	major NCD	China	71.8 (6.71)	0.94 [0.79, 0.99]	1.00 [0.90, 1.00]		
Liew 2015a	31	1	2	33	22/23	≥ secondary	major NCD	China	71.8 (6.71)	0.94 [0.80, 0.99]	0.97 [0.85, 1.00]		
Lu 2011	427	1100	14	5183	13/14; 19/20; 24/25	0; 1 to 6; ≥7	Mixed	China	75.1 (0.9)	0.97 [0.95, 0.98]	0.82 [0.82, 0.83]		
Tan 2015	260	135	25	546	24	6 (6)	Mixed	China	60-79	0.91 [0.87, 0.94]	0.80 [0.77, 0.83]		
Tan 2015a	501	267	114	1200	21	6 (6)	Mixed	China	80-89	0.81 [0.78, 0.84]	0.82 [0.80, 0.84]		
Tan 2015b	19	14	4	41	19	6 (6)	Mixed	China	≥90	0.83 [0.61, 0.95]	0.75 [0.61, 0.85]		
Tsai 2012	96	2	2	36	22	8.4 (4.9)	AD	China	79.2 (6.8),	0.98 [0.93, 1.00]	0.95 [0.82, 0.99]		
Yeung 2014	45	11	4	119	19	3.26 (4.03)	AD	China	76.45 (7.54)	0.92 [0.80, 0.98]	0.92 [0.85, 0.96]		
Costa 2012	29	3	1	97	22	10.60 (2.42)	AD	Germany	67.80 (8.13)	0.97 [0.83, 1.00]	0.97 [0.91, 0.99]		
Fujiwara 2010	30	4	0	32	26	12.1 (3.0)	AD	Japan	77.3 (6.3)	1.00 [0.88, 1.00]	0.89 [0.74, 0.97]		
Kasai 2011	18	42	6	56	22	8	MCI+	Japan	82.6	0.75 [0.53, 0.90]	0.57 [0.47, 0.67]		
Lee 2008	43	12	1	104	21/22	7.9 (3.7)	AD	Korea	70.4 (8.6)	0.98 [0.88, 1.00]	0.90 [0.83, 0.95]		
Dominguez 2014	81	141	16	368	21	7.54 (4.33)	AD	Phillipines	74.47 (6.35)	0.84 [0.75, 0.90]	0.72 [0.68, 0.76]		
Freitas 2012	39	1	11	49	17	6.18 (3.71)	BV_FTD	Portugal	67.96 (7.69)	0.78 [0.64, 0.88]	0.98 [0.89, 1.00]		
Freitas 2012a	26	1	8	33	17	4.97 (2.75)	VaD	Portugal	73.21 (7.85)	0.76 [0.59, 0.89]	0.97 [0.85, 1.00]		
Freitas 2013	79	4	11	176	17	6.23 (4.12)	AD	Portugal	70.52 (7.95)	0.88 [0.79, 0.94]	0.98 [0.94, 0.99]		
Karunaratne 2011	48	10	1	39	24	9.54 (3.1)	AD	Sri Lanka	70.45 (10.19)	0.98 [0.89, 1.00]	0.80 [0.66, 0.90]		
Wang 2013	64	11	3	51	21/22	6.42 (4.72)	AD	Taiwan	78.19 (5.71)	0.96 [0.87, 0.99]	0.82 [0.70, 0.91]		
Wang 2013a	33	12	3	50	22/23	7 (5.1)	DLB	Taiwan	77.31 (6.17)	0.92 [0.78, 0.98]	0.81 [0.69, 0.90]		

Figure 5: Forest plot for studies differentiating between dementia and NC.

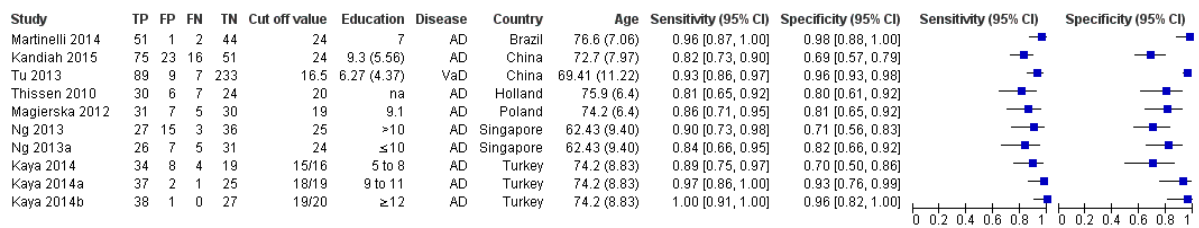| Study	TP	FP	FN	TN	Cut off value	Education	Disease	Country	Age	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Martinelli 2014	51	1	2	44	24	7	AD	Brazil	76.6 (7.06)	0.96 [0.87, 1.00]	0.98 [0.88, 1.00]		
Kandiah 2015	75	23	16	51	24	9.3 (5.56)	AD	China	72.7 (7.97)	0.82 [0.73, 0.90]	0.69 [0.57, 0.79]		
Tu 2013	89	9	7	233	16.5	6.27 (4.37)	VaD	China	69.41 (11.22)	0.93 [0.86, 0.97]	0.96 [0.93, 0.98]		
Thissen 2010	30	6	7	24	20	na	AD	Holland	75.9 (6.4)	0.81 [0.65, 0.92]	0.80 [0.61, 0.92]		
Magierska 2012	31	7	5	30	19	9.1	AD	Poland	74.2 (6.4)	0.86 [0.71, 0.95]	0.81 [0.65, 0.92]		
Ng 2013	27	15	3	36	25	>10	AD	Singapore	62.43 (9.40)	0.90 [0.73, 0.98]	0.71 [0.56, 0.83]		
Ng 2013a	26	7	5	31	24	≤10	AD	Singapore	62.43 (9.40)	0.84 [0.66, 0.95]	0.82 [0.66, 0.92]		
Kaya 2014	34	8	4	19	15/16	5 to 8	AD	Turkey	74.2 (8.83)	0.89 [0.75, 0.97]	0.70 [0.50, 0.86]		
Kaya 2014a	37	2	1	25	18/19	9 to 11	AD	Turkey	74.2 (8.83)	0.97 [0.86, 1.00]	0.93 [0.76, 0.99]		
Kaya 2014b	38	1	0	27	19/20	≥12	AD	Turkey	74.2 (8.83)	1.00 [0.91, 1.00]	0.96 [0.82, 1.00]		

Figure 6: Forest plot for studies differentiating between MCI and dementia.

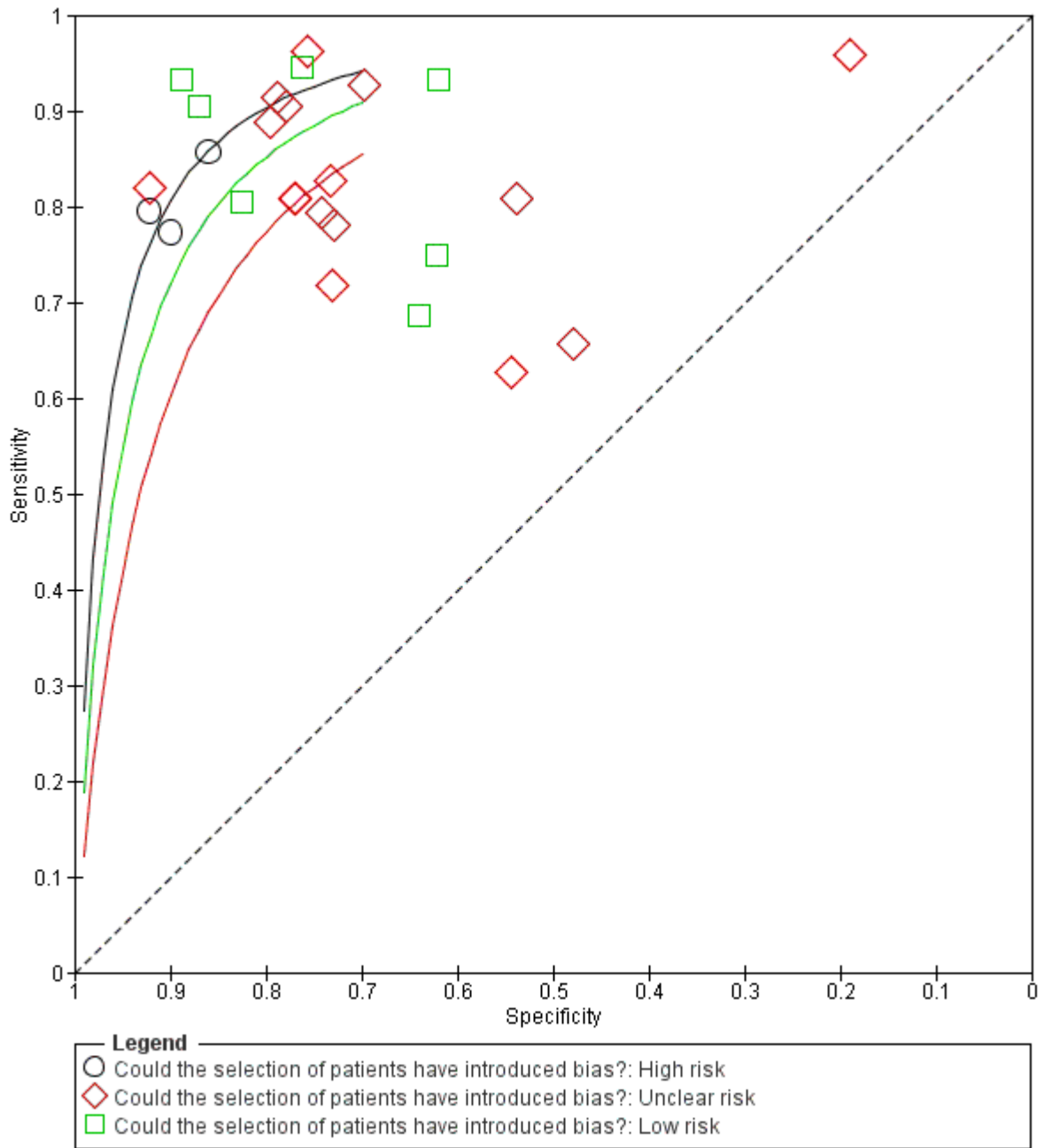Figure 7: *SROC curve displaying influence of quality item, selection of patients on accuracy estimates.*

Supplementary Material

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Chu 2015	High	Low	Low	Low	Low	Low	Low
Costa 2012	High	Low	Low	Low	Low	Low	Low
Dong 2013	High	Low	Low	Low	Unclear	Low	Low
Freitas 2013	High	Unclear	Low	Low	Low	Low	Low
Fujiwara 2010	High	Low	Low	Low	Low	Low	Low
Gil 2015	Unclear	Low	Low	Low	Low	Low	Low
Hu 2013	High	Low	Low	Low	Low	Low	Low
Julayanont 2015	High	Low	Low	Low	Low	Low	Low
Kaya 2014	High	Unclear	Unclear	Low	Low	Low	Low
Lee 2008	High	Unclear	Low	Low	Low	Unclear	Low
Liew 2015	High	Unclear	Low	Unclear	Low	Low	Low
Lifshitz 2012	Low	Unclear	Unclear	Low	Low	Unclear	Unclear
Lu 2011	Low	Unclear	Low	Low	Low	Low	Low
Magierska 2012	High	Low	Low	Low	Low	Low	Low
Martinelli 2014	Unclear	Unclear	Low	Low	Low	Unclear	Low
Memoria 2013	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Unclear
Munoz-Neira 2014	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Low
Ng 2008	High	Low	Low	Low	Low	Low	Low
Ng 2013	High	Low	Low	Low	Unclear	Low	Low
Ng 2015	High	Unclear	Low	Low	Unclear	Unclear	Low
Tan 2015	Low	Unclear	Low	Unclear	Low	Unclear	Low
Thissen 2010	High	Unclear	Low	Low	Unclear	Unclear	Low
Tsai 2012	High	Unclear	Unclear	Low	Low	Unclear	Unclear
Tu 2013	High	Low	Low	Low	Unclear	Low	Low
Yeung 2014	High	Unclear	Low	Low	Unclear	Low	Low
Yu 2012	Low	Unclear	Low	Low	Low	Low	Low
Zhao 2011	High	Unclear	Low	Unclear	Unclear	Low	Low
Zhou 2014	Low	Unclear	Unclear	Low	Low	Unclear	Low

Legend: — High   ? Unclear   + Low

Figure 1. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study MCI VS NC

Figure 1. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study DEMENTIA VS NC

Figure 1. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study MCI VS DEMENTIA