Journal of Alzheimer's Disease Reports 7 (2023) 973–987
DOI 10.3233/ADR-230024
IOS Press

Systematic Review

Systematic Review and Meta-Analysis of Brief Cognitive Instruments to Evaluate Suspected Dementia in Chinese-Speaking Populations

Ruan-Ching Yu[a], Jen-Chieh Lai[b], Esther K. Hui[a], Naaheed Mukadam[a], Narinder Kapur[c], Joshua Stott[c] and Gill Livingston[a,*]
[a]*Department of Mental Health of Older People, University College London, London, UK*
[b]*Department of Neurology, Hualien Tzu Chi Medical Centre, Hualien, Taiwan*
[c]*Department of Clinical, Educational and Health Psychology, University College London, London, UK*

Received 20 March 2023
Accepted 7 August 2023
Pre-press 25 August 2023
Published 1 September 2023

Abstract.
Background: Chinese is the most commonly spoken world language; however, most cognitive tests were developed and validated in the West. It is essential to find out which tests are valid and practical in Chinese speaking people with suspected dementia.
Objective: We therefore conducted a systematic review and meta-analysis of brief cognitive tests adapted for Chinese-speaking populations in people presenting for assessment of suspected dementia.
Methods: We searched electronic databases for studies reporting brief ($\leq$20 minutes) cognitive test's sensitivity and specificity as part of dementia diagnosis for Chinese-speaking populations in clinical settings. We assessed quality using Centre for Evidence Based Medicine (CEBM) criteria and translation and cultural adaptation using the Manchester Translation Reporting Questionnaire (MTRQ), and Manchester Cultural Adaptation Reporting Questionnaire (MCAR). We assessed heterogeneity and combined sensitivity in meta-analyses.
Results: 38 studies met inclusion criteria and 22 were included in meta-analyses. None met the highest CEBM criteria. Five studies met the highest criteria of MTRQ and MCAR. In meta-analyses of studies with acceptable heterogeneity ($I^2 < 75\%$), Addenbrooke's Cognitive Examination Revised &III (ACE-R & ACE-III) had the best sensitivity and specificity; specifically, for dementia (93.5% & 85.6%) and mild cognitive impairment (81.4% & 76.7%).
Conclusions: Current evidence is that the ACE-R and ACE-III are the best brief cognitive assessments for dementia and mild cognitive impairment in Chinese-speaking populations. They may improve time taken to diagnosis, allowing people to access interventions and future planning.

Keywords: Alzheimer's disease, cognitive assessment, dementia, diagnosis, mild cognitive impairment

*Correspondence to: Gill Livingston, MD, Department of Mental Health of Older People, University College London, 6th Floor, Maple House 149 Tottenham Court Road, London W1T 7NF, UK. Tel.: +44 20 7679 9435; E-mail: g.livingston@ucl.ac.uk.

ISSN 2542-4823 © 2023 – The authors. Published by IOS Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (CC BY-NC 4.0).

INTRODUCTION

There are over 55 million people living with dementia worldwide [1] of whom over 11 million people are from Chinese-speaking populations (approximately 10 million in China, 226,000 in Taiwan, 123,000 in Malaysia, 115,000 in Hong Kong, and 45,000 in Singapore) [2]. Chinese is the most commonly spoken language in the world, with a growing number of speakers (1.31 billion in 2020 [3]). It is the national language in The People's Republic of China (including Hong Kong and Macau), Taiwan, Malaysia, and Singapore. Chinese-speaking populations have a different education, ethnicity, and culture than Western countries, where most cognitive tests were initially validated and implemented [4–6]. Consequently, there is an urgent need to evaluate the performance of diagnostic tests that have been adapted for Chinese-speaking populations.

Cognitive assessments used routinely in clinical settings should be quick, and, accurate to assist with timely dementia and mild cognitive impairment (MCI) diagnosis [1]. Screening tests are conducted on a larger population, typically lacking symptoms but with a risk factor, such as age, and the aim is to decide who warrants further diagnostic investigation of an illness. Diagnostic tests are performed when a person is suspected of having a specific disease. The purpose of these tests is to confirm or refute the suspected diagnosis [7].

In recent years, systematic reviews focusing on brief cognitive assessments used in dementia have been completed for English [8] and Spanish [9] speaking populations. In addition, a more recent study among Chinese-speaking populations [6] focused on the tests administered in multiple contexts (e.g., primary care, secondary care, care home) for screening (not used as part of the diagnostic process), other types of measures (e.g., behavior tests), and with administration times up to 35 minutes. To our knowledge, there is currently no systematic review regarding the best-performing short cognitive Chinese diagnostic tools focusing specifically on people presenting to secondary care, which is important since such findings are of particular relevance to clinicians directly involved in dementia diagnosis.

This systematic review and meta-analysis about diagnostic test accuracy therefore aims to find the optimal brief cognitive assessment for people with suspected dementia or mild cognitive impairment presenting to secondary care (e.g., memory clinic, neurology department) among Chinese-speaking populations. We assessed the level and quality of evidence, and the quality of cultural adaptation and translation for each study, as people from different Chinese-speaking populations live in differing historical, economic, and societal contexts.

METHODS

This systematic review was registered on PROSPERO (registration number: CRD42019134092) and followed the standard guidelines for conduct and reporting systematic reviews of diagnostic studies, including Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [10], PRISMA-diagnostic test accuracy (PRISMA-DTA) checklist [11], and guidelines from the Cochrane Diagnostic Test Accuracy Working Group [12].

Data sources and search strategy

We searched the following electronic databases: Embase, Ovid MEDLINE®, PsychInfo, PsycTESTS, Web of Science core collection, the Cochrane library (Cochrane library of Systematic Reviews, Cochrane Central Register of Controlled Trials (CENTRAL) and Cochrane Methodology Register) from inception to October 12, 2021.

We searched abstract, keywords, and titles using the following search terms (see the Supplementary Material for details of the research protocol):

"Chinese" or "Mandarin" or "Hokkien" or "Hoklo" or "Cantonese" or "Hakka" or "Taiwan*" or "China" or "Hong Kong*" or "Singapore*" or "Macao" or "Malaysia.

And "Alzheimer*" or "AD" or "dement*" or "VaD" or "FTD" or "Mild cognitive impairment" or "MCI" or "memory loss".

And (all fields) "assessment" or "evaluation" or "scale" or "test" or "tool" or "Instrument" or "battery" or "measure*" or "screen*" or "diagnos*" or "inventory*" or "validat*"

Inclusion criteria

Studies had to fulfil all the following criteria:

- Validation studies of instruments to assess cognition as part of diagnosis (not for the purpose of screening) for patients with any suspected dementia or MCI in a memory clinic or similar setting where people present with suspected dementia;

- In Chinese-speaking populations;
- Time taken $\leq$20 minutes (either stated in paper or in other publications);
- Patient assessed face-to-face;
- Published in peer reviewed journals;
- Published in English or Chinese.

Exclusion criteria

We excluded instruments testing function or behavior; telephonic or computerized self-administration tests, informant questionnaires; instruments detecting dementia praecox or dementia secondary to head injury; studies in people diseases other than dementia and MCI; those whose purpose was to stage severity of rather than test for suspected dementia or MCI; learning disability populations and qualitative studies. We also excluded studies which only reported qualitative results, or did not report cut-off scores, sensitivities, and specificities; and those in which the instruments were used as a screening tool not as part of the diagnostic processes in people presenting with suspected dementia.

Study selection

We exported all searches to Endnote and removed duplicates. One author (R-CY) screened the titles and abstract of retrieved articles and excluded irrelevant papers. Then two authors (R-CY, J-CL) independently assessed the remaining full-texts for inclusion. Disagreements about inclusion were resolved by discussion with other authors until a consensus was reached.

Data extraction and definition

Two authors (R-CY, J-CL) independently extracted data from each included paper using a standard form. We recorded data on population, recruitment settings, specification of illness, study design, purpose of the test, time taken to administer test, total items, total scores, cut-off score, sensitivity, specificity, validity, reference standard, and blinding. A consensus was reached in case of any disagreements.

Quality assessment and evaluation of translation and cultural adaptation procedures

Two of the three authors independently assessed the quality of studies (R-CY: all studies, J-CL: studies with clinical controls, E-H: studies with community controls) and the quality of the reported translation and cultural adaptation procedures using the Centre for Evidence-based Medicine (CEBM) diagnostics criteria [13] and Manchester Translation Reporting Questionnaire (MTRQ) and Manchester Cultural Adaptation Reporting Questionnaire (MCAR) [14]. respectively. Subsequently, GL or NM reviewed inconsistent scores and the team determined the final score for each item in all scales after discussing and reaching a consensus. We used the Kappa statistic for calculating interrater reliability for MTRQ and MCAR.

Compared to other tools that are more complex and domain-specific (e.g., Quality Assessment of Diagnostic Accuracy Studies (QUADAS)), CEBM is simple to use, can be applied to a wide range of diagnostic accuracy studies, and evaluates multiple aspects of study quality, including study design, methodology, and reporting. The total possible score was 10 with higher scores indicating higher quality. Criteria which were not fulfilled or specified were scored as zero. Fully met criteria were scored as two, we did not score 'partially fulfilled' as one, as has been done before [13] because we found it difficult to distinguish partially fulfilled from not fulfilled. The scoring criteria and studies scoring are in the Supplementary Material. MTRQ and MCAR were specifically developed to rate the quality of translation and cultural adaptation procedures, with seven items in each scale rating the quality of reported procedures (see Supplementary Table 1). When the cultural adaptation or translation procedures were not mentioned, the studies received an adaptation score of zero. If they were mentioned with no details, a score of one was given. To score two, they were mentioned with insufficient details for replication (2a) or referred to another publication (mainly original authors of the test) which provided insufficient details for replication (2b). When pre-existing guidelines about adaptation were provided, but with insufficient details for replication, it scored three. To score four, either the adaptation was mentioned with sufficient details for replication (4a) or referred to another publication which provides sufficient details for replication (4b).

Data synthesis and statistical analysis

We conducted a meta-analysis if cognitive tests results were reported in more than one study. As dementia and MCI are different with many people with MCI never developing dementia, a diagnostic

test might be useful for one and not for the other. We therefore separated these results into dementia and MCI categories. We also separately reported the cut-off scores based on where the non-dementia groups (clinical controls versus community-based controls) were recruited as the populations and therefore cut-off scores may vary. The statistical methods used are described below:

Forest plots for combined sensitivity, specificity, and heterogeneity

We used R software, version 4.1, and used a univariate random-effects model to analyze the sensitivity and specificity of each brief cognitive test separately. Through applying the package "meta" and function "metaprop", the logit transformation and Clopper–Pearson method were used. First, the total effect size using the number of events and the number of samples from proportion-type data was calculated. Second, forest plots were used to graphically present the combined sensitivity and specificity. As 75% of sensitivity or specificity is not good enough to rule out an illness and the sum of sensitivity and specificity should be at least 150% [15], we therefore defined either sensitivity or specificity of individual papers of meta-analyzed tests below 75% as unacceptable, 75–90% as satisfactory, and higher than 90% as excellent. Third, statistical heterogeneity among the trials was calculated as I^2, which describes the percentage of total variation across studies due to the heterogeneity. An I^2 value 0% to 40% might not be important, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity, with over 75% used as a cut-off to indicate considerable heterogeneity [16].

In addition, to graphically display how the summary sensitivity changes with the summary specificity and vice versa, we generated hierarchical summary receiver operating characteristic (HSROC) and summary receiver operating characteristic (SROC) plots using the command of "metandi" and "metandiplot" on Stata, and "reitsma" on R software (when the number of studies was <3).

Sensitivity analysis on univariate and bivariate models

As there is debate in the literature as to the best method for meta-analyzing data on diagnostic test accuracy, we conducted a sensitivity analysis in Stata/MP 17.0 to compare univariate and bivariate random-effects methods using the command of "metadta" [17].

Meta-regression

We conducted a random effects meta-regression on Stata/MP 17.0 to consider possible causes for heterogeneity (population, subtype, reference standard, scoring system), with the command "gsort" [18].

Publication bias

We used Deeks' funnel plot asymmetry test [19] on R software for investigating publication bias, as it was developed for use in diagnostic test accuracy meta-analyses [20]. This entailed conducting a regression of diagnostic log odds ratio against 1/sqrt (effective sample size), with $p < 0.05$ for the slope coefficient indicating significant asymmetry.

RESULTS

Study selection

We retrieved 19,723 references, of which 38 studies met inclusion criteria (see PRISMA diagram Fig. 1 [10] and Supplementary Table 2 for PRISMA-DTA checklist [11]). Among all studies that met the inclusion criteria, 22 studies were analyzed in the meta-analysis, 18 studies that used other brief cognitive assessments in a single Chinese-speaking population were included a narrative review. All 38 studies underwent translation and cultural adaptation procedures.

Inter-rater reliability of the MTRQ and MCAR

The inter-rater (R-CY, J-CL) reliability coefficient for MTRQ and MCAR were 0.835 and 0.907 respectively.

Study characteristics and quality analysis

Table 1 presents the included articles, instrument characteristics, the study quality and the translation and cultural adaptation quality. The 16 studies that recruited participants only from clinical settings were from China (8), Taiwan (5), Hong Kong (2), and Singapore (1). The 22 studies that recruited controls from community settings were from China (12), Taiwan (5), Singapore (3), and Hong Kong (2). Studies are

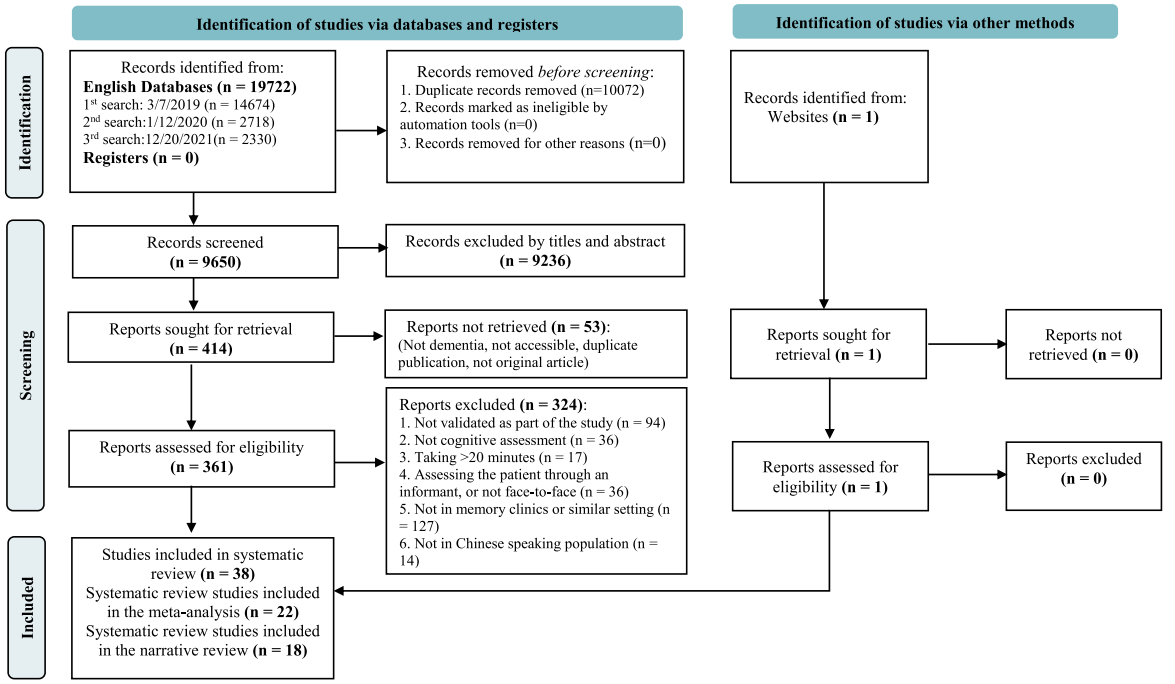


Fig. 1. PRISMA diagram.

presented by settings and further grouped by country and cognitive test.

None met the highest score of CEBM: two studies recruited participants through different standard measurements [21, 22] and no information was provided for the gold standard [23, 24] or not blinded to other cognitive tests score [25, 26]. Three studies did not specify whether they were blinded or not [21, 27, 28]. 14 studies provided insufficient details on the assessments [21, 23–25, 29–37]. Only one study [38] validated the diagnostic test in a representative spectrum of patients (like those in whom it would be used in practice). All others excluded participants with specific conditions, such as eyesight problems, stroke, or psychiatric issues.

Five studies [26, 34, 39–41] achieved the highest score of 4 on the MTRQ and MCAR, twenty-seven studies scored 0 to 4, six studies scored 0 (see Supplementary Table 1).

Meta-analyses of instruments with multiple results

There were 22 studies included in meta-analysis: 16 reported populations with dementia and 14 with MCI. As some articles validated several tests, the number of studies is higher than the number of articles. The tests that could be meta-analyzed were:

- For dementia (7 tests): *Addenbrooke's Cognitive Examination III & Revised (ACE-III &-R), Clock Drawing Test (CDT), Mattis Dementia Rating Scale (DRS), Mini Mental State Examination (MMSE),* and *Montreal cognitive assessment (MoCA), Shape Trail Test-A, and B (STT-A&B)*
- For MCI (4 tests): *ACE-III &-R, MMSE, MoCA, Montreal cognitive assessment-Basic (MoCA-BC)*

Please see Supplementary Table 3 (dementia) and Supplementary Table 4 (MCI) for more comprehensive details (e.g., the gold-standard criteria used, number of participants, and the positive/negative likelihood ratio). Supplementary Table 5 shows a 2×2 table of true positive (TP), false negative (FN), true negative (TN), and false positive (FP). Supplementary Figure 1 demonstrates the overall performance of the eight tests through HSROC and SROC curves.

Diagnostic test accuracy, heterogeneity, and cut-off scores for dementia in meta-analysis

MoCA in meta-analysis had a sensitivity (I^2) of 91.9% (67%) and specificity (I^2) of 94.3% (22%) (see Fig. 2). Using an adjusted MoCA scoring system, the

Table 1
Brief cognitive tests for dementia and mild cognitive impairment and quality score (Arranged by country, setting, test name)

Settings, and Author	Test Code,		Test Name	Illness	Time taken in individual study (min)	Total Score	IRR	TRR	CEBM	MTRQ	MCAR		Analysis
A	C1	Li et al., 2021 [33]	FAB- Phonemic	AD, aMCI, naMCI	5	18	0.997	0.819	6	0	2a		Δ
	C2	Shi et al., 2012 [35]	HVLT	dementia, AD, aMCI	NA	36	NA	NA	6	0	0		Δ
	C3	Yang et al., 2019 [42]	[M-ACE], MMSE	mild dementia, MCI	<5	30	NA	NA	8	4a	2a	X	Δ
	C4	Li et al., 2018 [32]	Mini-Cog	MCI	3	9	NA	NA	4	0	0		Δ
	C5	Wen et al., 2008 [36]	MoCA	MCI	<10	30	NA	NA	4	0	4a	X	
	C6	Li, Jia, et al., 2018 [23]	MoCA	MCI	NA	30	NA	NA	4	0	0		
	C7	Xu et al., 2003 [43]	MMSE	dementia	15	NA	NA	0.75	8	2b	4a	X	
	C8	Wei et al., 2018 [44]	TMT (A & B)	dementia, MCI, VaMCI,	150/ 300 secs	NA	NA	NA	6	2a	0		Δ
B	C9	Wang et al., 2017 [45]	ACE-III	dementia	NA	100	NA	NA	8	1	4a	X	
	C10	Fang et al., 2014 [30]	ACE-R	mild AD, aMCI	NA	100	0.994	0.967	4	1	4a	X	
	C11	Guo et al., 2002 [46]	AFT	AD, aMCI	1	unlimited	NA	NA	4	0	2a		Δ
	C12	Guo et al., 1991 [47]	BNT	AD, aMCI	NA	30	NA	NA	4	0	2a		Δ
	C13	Qian et al., 2021 [48]	DRS-CV	AD, aMCI	NA	144	NA	NA	6	1	1		
	C14	Guo et al., 2004 [49]	MDRS	Mild AD	15	144, 124 (illiterate)	NA	NA	4	2b	2b	X	
	C15	Chen et al., 2016 [50]	MoCA-BC	AD, MCI	NA	30	0.96	NA	8	1	1	X	
	C16	Huang et al., 2018 [27]	MoCA-BC	AD, MCI	NA	30	NA	NA	6	2b	1	X	
	C17	Zhang et al., 2019 [28]	MoCA-BC	AD, MCI	NA	NA	NA	NA	4	2b	2b		
	C18	Guo et al., 2010 [51]	[QCST], MoCA, MMSE	mild AD, MCI	8–15	90	NA	0.93	6	0	4a	X	Δ

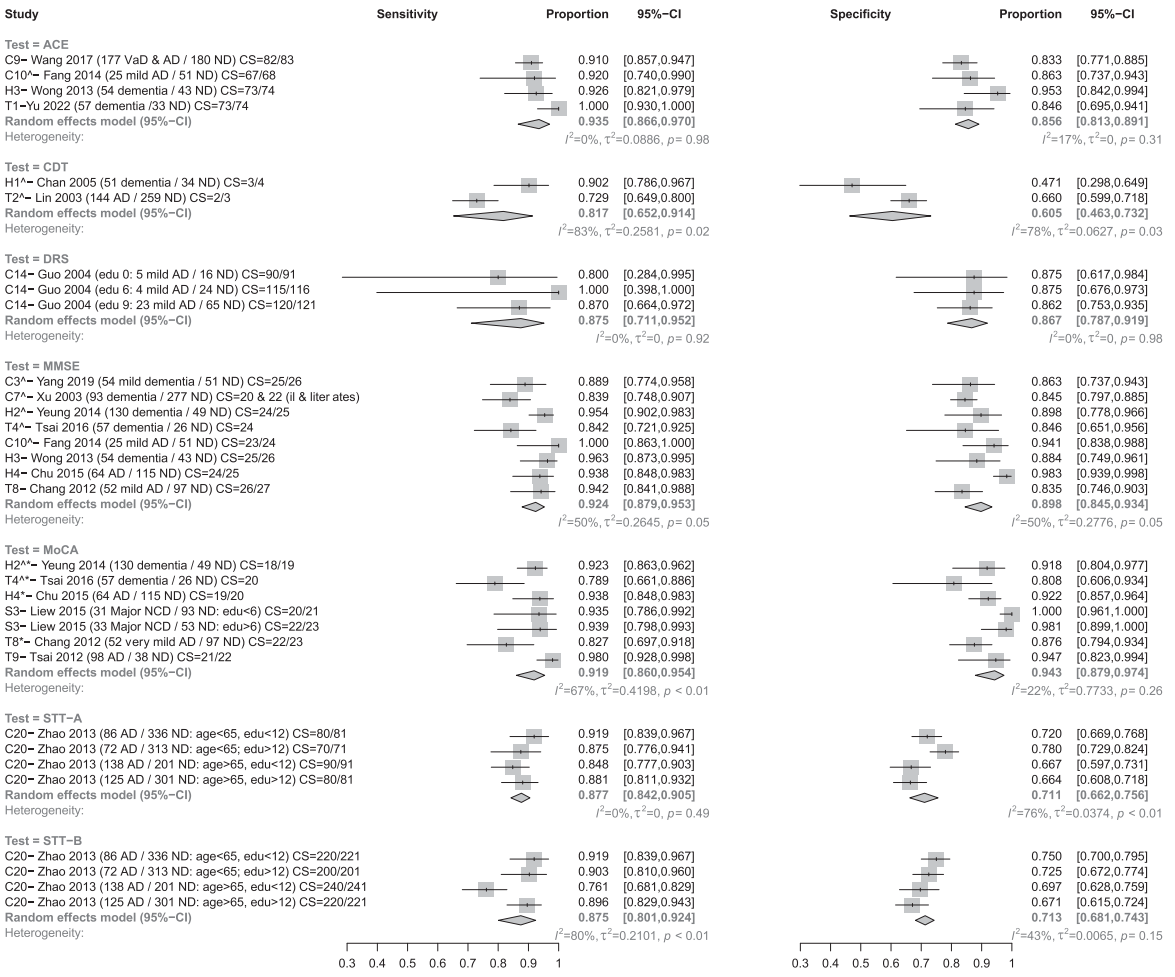
	C19	Huang et al., 2019 [52]	[silhouettes-A], MoCA, STT-B, & other tests	AD, MCI	NA	15	0.99	0.91–0.98	8	2a	2a	X	Δ
	C20	Zhao et al., 2013 [53]	STT(A&B)	AD, aMCI	around 5	NA	NA	NA	8	0	4a	X	
							Average		5.8	1.0	2.1		
A	H1	Chan et al., 2005 [54]	[CDT], [T&C]	dementia	CDT = 90.9, T&C = 65.6 secs	CDT = 10, T&C = pass or not	NA	NA	8	0/0	0/0	X	Δ
	H2	Yeung et al., 2014 [25]	[MoCA], MMSE	dementia, MCI	10–15	30	0.987	0.92	4	0	0	X	
B	H3	Wong et al., 2013 [37]	[ACE-R], MMSE	dementia, MCI	around 15	100	1	1	6	2a	4a	X	
	H4	Chu et al., 2015 [40]	[MoCA], MMSE	AD, aMCI	10	30	0.96	0.95	4	4a	4a	X	
							Average		5.5	1.2	1.6		
A	S1	Malhotra et al., 2013 [24]	SPMSQ	dementia, MCI	NA	10 (errors)	NA	NA	4	1	0		Δ
B	S2	Chong et al., 2010 [21]	[FAB]	dementia, MCI	5	NA	NA		2	2b	4b		Δ
	S3	Liew et al., 2015 [22]	MoCA	dementia	NA	NA	NA	NA	2	1	0		
	S4	Low et al., 2020 [34]	VCAT	AD, MCI	15.7	30	NA	NA	4	4a	4a	X	
							Average		3.0	2.0	2.0		Δ
A	T1	Yu et al., 2021 [26]	ACE-III	dementia	NC:16.3 Dementia:24.0	100	NA	NA	6	4a	4a	X	
	T2	Lin et al., 2003 [41]	CDT	very mild & mild AD	NA	3	0.99	0.88–0.9	6	4a	4a	X	
	T3	Chiu et al., 2008 [55]	CDT	very mild & mild AD, MCI	NA	11/10	0.83/0.87	NA	8	0	0		Δ

(*Continued*)

Table 1
(*Continued*)

Settings, and Author	Test Code,		Test Name	Illness	Time taken in individual study (min)	Total Score	IRR	TRR	CEBM	MTRQ	MCAR	Analysis	
	T4	Tsai et al., 2016 [56]	[MoCA], [MMSE]	dementia, MCI	15, 5–10	30, 30	NA	0.92	6	2b/0	4b/0	X	
	T5	Chen et al., 2015 [29]	RUDAS	AD	10	30	0.88	0.9	6	3	3		Δ
B	T6	Tsai et al., 2018 [57]	BHT-Cog part	dementia, MCI	4	16	NA	NA	4	N/A	N/A		Δ
	T7	Chang et al., 2010 [38]	CVVLT	AD	NA	NA	NA	NA	8	4a	4a		Δ
	T8	Chang et al., 2012 [39]	[MoCA], MMSE	AD	NA	30	NA	NA	6	0	4a	X	
	T9	Tsai et al., 2012 [58]	MoCA	AD, MCI	10	30	0.88	0.88	6	2a	4a	X	
	T10	Lee et al., 2018 [31]	[Qmci], MoCA, MMSE	dementia, MCI	<5	100	0.87	1.00	6	3	3	X	Δ
							Average		6.2	2.2	3.0		

The codes starting with syllables of C, H, S, T refer to China, Hong Kong, Singapore, and Taiwan, respectively; NA, Not available (the information is not provided in the study); The setting of A/ B refers to clinical or community-based controls, respectively; The test in bracket is the main validated test; * aside MoCA refers to the adjusted scoring system; X and Δ refer to the studies included in the meta-analysis and the narrative review, respectively; MCI, mild cognitive impairment; CEBM, the Centre for Evidence-based Medicine diagnostics criteria. The tests in the brackets are the target tests to be validated. ACE-R, Addenbrooke's Cognitive Examination Revised; AFT, Animal Fluency Test; BHT-cog, Brain Health Test-Cog part; BNT, Boston Naming Test; CDT, Clock Drawing Test; CFT-C, Rey-Osterrieth Complex Figure Test-Copy; CVVLT, Chinese version of the Verbal Learning Test; DRS/MDRS, Mattis dementia rating scale; FAB, Frontal Assessment Battery; HVLT, Hopkins Verbal Learning Test; JLO, Judgment of Line Orientation; M-ACE, Mini-Addenbrooke's Cognitive Examination; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MoCA-BC, Montreal Cognitive Assessment Basic; Qmci, Quick Mild Cognitive Impairment Screen; RUDAS, Rowland Universal Dementia Assessment Scale; SPMSQ, Short Portable Mental Status Questionnaire; STT, Shape Trail Test; T&C, Time and Change Test; TMT, Trail-Making Test; VCAT, Visual Cognitive Assessment Test. IRR, inter-rater reliability; TRR, test-retest reliability; CEBM, Centre for Evidence-Based Medicine.



The codes starting with syllables of C, H, S, T refer to China, Hong Kong, Singapore, and Taiwan, respectively. CI = confidence intervals; CS = cut-off scores; edu = years of education; ^indicates the controls were recruited from the clinics; * Indicates that the MoCA applied different scoring system; ND = non-dementia; AD = Alzheimer's disease; VaD = vascular dementia; NCD = neuro cognitive disorder; ACE = Addenbrooke's Cognitive Examination III & Revised; CDT = Clock Drawing Test; DRS = Mattis Dementia Rating Scale; MMSE = Mini-Mental State Examination; MoCA = Montreal cognitive assessment; STT-A&B = Shape Trail Test-A, and B.

Fig. 2. Meta-analyses for diagnostic test accuracy on diagnosing dementia.

best cut-off scores for separating dementia from clinical controls were 18/19 in Hong Kong [25] and 20 in Taiwan [59]. When the controls were selected from the community, the best Alzheimer's disease (AD) cut-off score in Hong Kong was 19/20 [40]. Using non-adjusted scoring system, the suggested AD cut-off scores in Taiwan were 21/22 [58], as well as 20/21 (years of education <6) and 22/23 (years of education >6) for major neurocognitive disorder (NCD) in Singapore [22].

MMSE' sensitivity (I^2) and specificity (I^2) in meta-analysis were 92.4% (50%) and 89.8% (50%), respectively. When the controls were recruited from the clinics, the suitable cut-off scores were 24/25 for dementia in Hong Kong [25], 22 and 20 for dementia in the population with or without education in China [43], and 23/24 and 25/26 for AD [30] and mild dementia [42] in China. Compared to the community-based controls, the optimal cut-off score was 26/27 for mild AD in Taiwan [39], and 24/25 for AD [40] and 25/26 for all-cause dementia [37] in Hong Kong.

ACE (ACE-III & ACE-R)'s sensitivity (I^2) and specificity (I^2) in meta-analysis were 93.5% (0%) and 85.6% (17%). For studies with clinical controls,

the appropriate ACE-R cut-off score for mild AD was 67/68 in China [30] and 73/73 for dementia in Taiwan [26]. For studies with community-based controls, ACE-III and ACE-R's best cut-off scores were 83 for VaD and AD in China [45] and 73/74 for dementia in Hong Kong [37], respectively.

DRS's sensitivity (I^2) and specificity (I^2) in meta-analysis were 87.5% (0%) and 86.7% (0%) in identifying dementia from community-based controls, with cut-off scores of 90 and 120 for AD with zero and nine year(s) of education in China [49], respectively.

STT-A and STT-B's, the sensitivity (I^2) and specificity (I^2) in meta-analysis were 87.7% (0%) and 71.1% (76%), 87.5% (80%) and 71.3% (43%) in identifying dementia from community-based controls, respectively. For AD, STT-A and -B best cut-off scores were 80 and 220 for those aged <65 with <12 years of education, 70 and 200 for those aged <65 with >12 years of education, 90 and 240 for those aged >65 with <12 years of education, 80 and 220 for those aged >65 with >12 years of education in China [53]

CDT's combined sensitivity (I^2) and specificity (I^2) were 81.7% (83%) and 60.5% (78%) in identifying dementia from clinical controls, with optimal cut-off scores of 3/4 for dementia in Hong Kong [54] and 2/3 for (AD) in Taiwan [41]

Diagnostic test accuracy and cut-off scores for MCI

Presented in Fig. 3, ACE (ACE-R)'s combined sensitivity (I^2) and specificity (I^2) were 81.4% (68%) and 76.7% (54%), with cut-off scores of 85/86 for separating aMCI from clinical controls in China [30], and 79/80 for differentiating MCI from community-based controls in Hong Kong [37].

MoCA-BC's combined sensitivity (I^2) and specificity (I^2) were 86.6% (81%) and 82.3% (85%), with cut-off scores of 19, 22, 24 for detecting MCI with low (<7), moderate (7–12), and high (>12) education levels from community-based controls in China, respectively [50].

MoCA's combined sensitivity (I^2) and specificity (I^2) were 82.7% (75%) and 79.6% (81%) [22]. With adjusted scoring systems, the suitable cut-off scores for distinguishing MCI from clinical controls were 25/26, 21/22, 24 in China [36], Hong Kong [25], and Taiwan [59]. When it comes to community-based controls, the best cut-off scores using original systems were 24, 23/24, and 22/23 for MCI in China [52], Taiwan [58], and Singapore [22], respectively. With adjusted scoring systems, 27/28 for aMCI in Hong Kong [40] and 23/24 for MCI in Taiwan [31] were recommended.

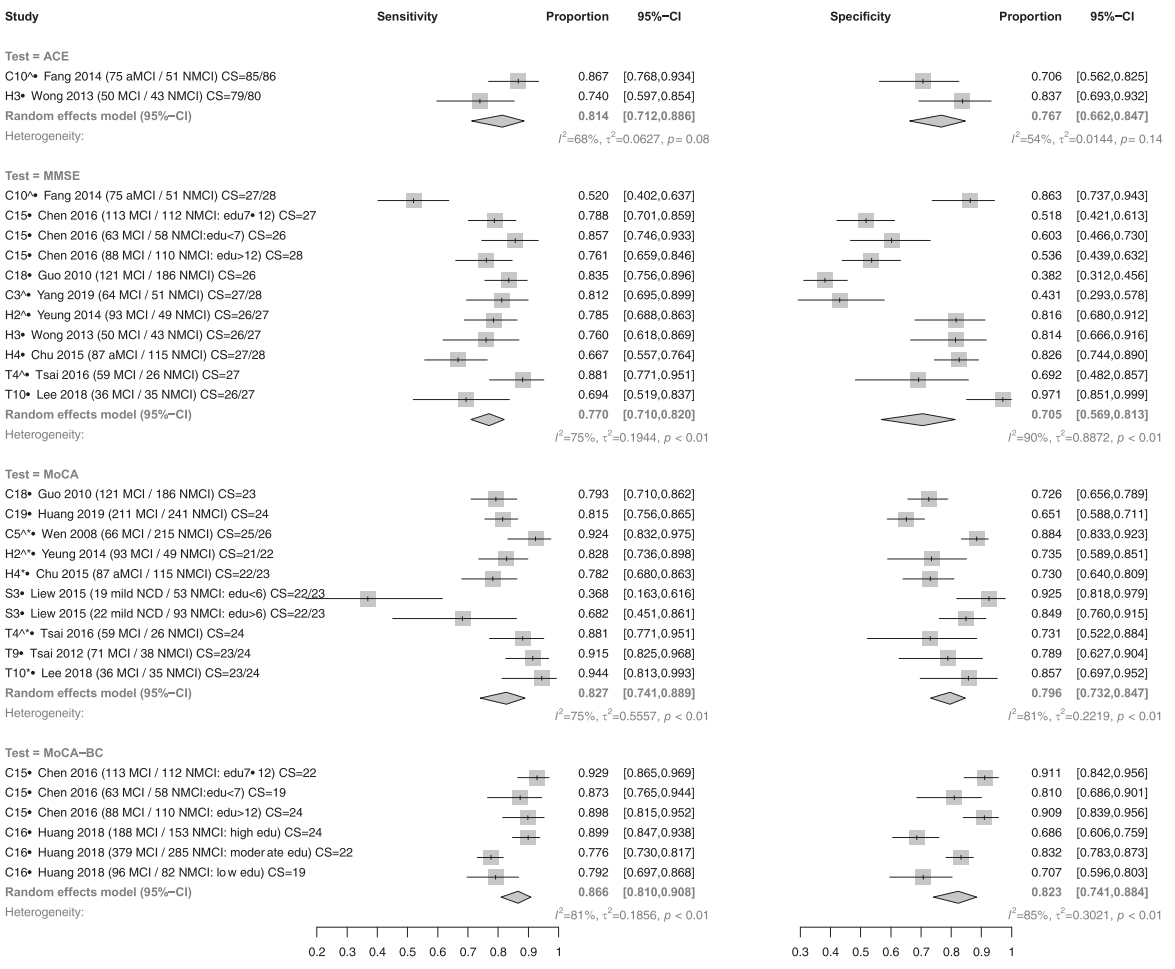
MMSE's combined sensitivity (I^2) and specificity (I^2) were 77.0% (75%) and 70.5% (90%), with best cut-off scores of 27/28 [42] and 22/23 [30] to separate MCI and aMCI from clinical controls in China, and 26/27 for MCI in Hong Kong [25]. The suggested cut-off scores were 26 in China [51], and 26/27 to differentiate MCI from community-based controls in Hong Kong [37] and Taiwan [31], and 27/28 for aMCI in Hong Kong [40].

Sensitivity analysis of univariate and bivariate analysis

Supplementary Table 6 shows the random-effect bivariate model. The confidence interval at 95% and the combined sensitivity and specificity were similar to the random-effect univariate model. In terms of heterogeneity, the random-effect bivariate model produces the generalized I^2 and Tau^2 values, which summarize the overall heterogeneity, by taking into account the correlation between sensitivity and specificity. Some generalized I^2 and Tau^2 values were less than those of univariate analytic models. For instance, the I^2 values of sensitivity and specificity were 50% and 50% in MMSE, 67% and 22% in MoCA for dementia, whereas it was 30% in MMSE and 0% in MoCA for the generalized model. The I^2 values of specificity in STT-A and that of sensitivity in STT-B were 76% and 80% in dementia patients, respectively, but 0% and 33% in the generalized model. However, these differences did not affect the findings on the best-performing tests.

Meta-regression

The random-effect meta-regression was performed on three tests (ACE, MMSE, and MoCA) to determine if the absolute sensitivity and specificity of each test differed by adding different covariates. The MMSE's specificity was significantly higher in other Chinese-speaking populations than in China (88% versus 70%), MMSE's sensitivity was significantly higher in MCI than other subtypes (80% versus 60%), and MoCA's original scoring system had significantly higher sensitivity (96% versus 88%) and specificity (98% versus 89%) than the modified scoring system



The codes starting with syllables of C, H, S, T refer to China, Hong Kong, Singapore, and Taiwan, respectively. CI, confidence intervals; CS, cut-off scores; edu, years of education; ^indicates the controls were recruited from the clinics; *indicates that the MoCA applied different scoring system; NMCI, non-MCI; MCI, mild cognitive impairment; aMCI, amnestic MCI; amMCI, aMCI-multiple domains; naMCI, nonamnestic MCI; NCD, neurocognitive disorder; ACE, Addenbrooke's Cognitive Examination III & Revised; MMSE, Mini-Mental State Examination; MoCA, Montreal cognitive assessment; MoCA-BC, Montreal cognitive assessment-Basic.

Fig. 3. Meta-analyses for diagnostic test accuracy on diagnosing MCI.

in the detection of dementia. The remaining factors had no impact on any of the tests.

Publication bias

For the meta-analysis of diagnostic accuracy tests (Supplementary Figure 2), the *p*-values for testing the asymmetry of the Deeks' funnel plots were 0.0030 and 0.6245 for patients with dementia and MCI, respectively. The results suggested that publication bias exists in validation of cognitive instruments in dementia but not in MCI populations.

Narrative review of single studies

The individual tests that were excluded in the meta-analysis (only appeared in one Chinese-speaking population), including the cognitive domains covered, administration time, and performances are discussed in Supplementary Table 12. The purpose is to sum up and convey results narratively when statistical or other formal methods of data estimation are not feasible, so as to gain a comprehensive understanding of the brief cognitive assessment instruments used in Chinese-speaking populations to detect dementia.

For other relevant information, please refer to Table 1. It lists the name of the tests, administration time, total score, inter-rater reliability (IRR), test-retest reliability (TRR), CEBM scores, translation (MTRQ) and cultural adaptation (MCAR) scores.

DISCUSSION

This is to our knowledge the first systematic review and meta-analysis of validation of brief cognitive tests used in secondary care for the diagnosis of dementia and MCI in Chinese speaking populations. The meta-analysis found that ACE (ACE-R & ACE-III) had the best validity (sensitivity (I^2) and specificity (I^2): 93.5% (0%) and 85.6% (17%) in the diagnosis of dementia; 81.4% (68%) and 76.7% (54%) in MCI). We also found that MMSE (13 studies) is the most tested brief cognitive test. None of the studies included met the highest standard of CEBM. Only five studies stated in detail how they translated and culturally adapted the tests into their own settings [26, 34, 39–41].

Among the seven tests for detecting dementia, ACE, DRS, MoCA, and MMSE all performed satisfactorily in detecting dementia (sensitivity and specificity >75%, heterogeneity <75%), while CDT, STT-A & B performed at an unacceptably low level (either sensitivity or specificity was lower than 75%).

When detecting MCI, MoCA-BC had the highest sensitivity and specificity among the three tests (MoCA-BC, MoCA, and ACE). However, when comparing the heterogeneity among those tests, only the ACE had acceptable I^2 values in both dementia and MCI, independent of statistical methods (univariate or bivariate model). For dementia, the I^2 values of STT-A & B were more than 75% in the univariate model; however, they were lower than 75% in the bivariate model. For MCI, the I^2 values of MMSE, MoCA, and MoCA-BC, the I^2 values were likewise higher than 75% in the univariate model but below 75% in the bivariate model (69%, 72%, and 74%) although they remained relatively high. The differences did not affect the conclusion that ACE is the test with satisfactory sensitivity and specificity and acceptable heterogeneity among individuals with dementia and MCI. In keeping with earlier findings, univariate and bivariate meta-analyses produced similar summary estimates and confidence intervals for sensitivity and specificity, and it may be that the difference between these approaches is insufficient to impact clinical decision-making [60]. Due to the low number of studies in ACE ($n = 2$) for MCI, as well as CDT ($n = 2$) and DRS ($n = 3$) for dementia, the uncertainty in the I^2 value must be acknowledged, despite the fact that ACE is the best-performing test. Systematic review studies that included patients from a variety of care settings, such as primary and secondary care, day care centers, and communities, concluded that the ACE-R may be useful for differentiating between different forms of dementia and should be administered when the diagnosis is uncertain [6, 8].

Similar to my findings, the previous study that used MTRQ and MCAR for the ACE, ACE-R, and ACE-III tests found that only seven out of 32 papers met the highest standard on both scales [14].

The result of the meta-regression found that the factors of population (China versus other areas), MCI subtype (MCI versus others), and scoring systems (original versus adjusted) contributed to some level of heterogeneity of MMSE and MoCA test. Other factors such as the choice of cut-off scores, educational levels of participants could not be fully examined due to the limited studies included in the meta-analysis. Future research with a larger sample size should evaluate those factors that may influence the diagnostic test accuracy for detecting dementia and MCI in Chinese-speaking-populations.

Publication bias was discovered in studies of individuals with dementia, but little is known about the mechanisms underlying publication bias in research of diagnostic test accuracy [61].

We used a thorough search strategy and robust assessment of quality and data extraction by three independent reviewers in line with recommendations. We included Chinese language papers identified in English databases. We compared the outcomes between univariate and bivariate random-effect methods to confirm our findings and found the same results. We examined possible reasons of high heterogeneity through meta-regression. Also, we included not only English language papers, but also Chinese language papers identified in English databases, in which there were four papers published in simplified Chinese from China. We had no restriction on the date of publication, which allowed this review to gather results thoroughly and enables future research to readily update our findings.

However, a limitation of the present study is that we did not search Chinese databases, this was for two main reasons. One is that the Chinese academic publishing system differs from western systems and faces unique challenges, such as the variable and contested

quality of Chinese journals (e.g., non-transparent editorial peer review and assessment processes), a lack of specialization and focus in newly developing and specialized fields, and the problems of low-quality papers (e.g., broker selling papers and researchers purchasing authorship) [62]. The systematic review and meta-analysis of cognitive instruments revealed that the published language was one of the factors contributing to the high heterogeneity of MoCA tests, with Chinese-language publications reporting much higher sensitivity and specificity [6]. To maintain a consistent level of quality throughout the electronic resources, we chose to include well-established and peer-reviewed English databases without language constraints, permitting the inclusion of high-quality Chinese publications. The Chinese databases, including CNKI and Airiti, may contain relevant research that can be incorporated into future studies, but we used journals which were all peer reviewed in this study.

We recommend future validation studies should fully report the translation and cultural adaptation procedures as it is essential when validating one cognitive assessment into different cultures. On current evidence we would recommend routinely using the ACE-R, ACE-III to assess people with suspected dementia in Chinese-speaking populations.

ACKNOWLEDGMENTS

The authors have no acknowledgments to report.

FUNDING

Ruan-Ching Yu is funded by the Ministry of Education in Taiwan. Esther K. Hui is funded by the UCL Overseas Research Scholarship and a UCL Doctoral School Fellowship. Naaheed Mukadam is funded by an Alzheimer's Society Senior Research Fellowship. Gill Livingston is supported By NIHR North Thames ARC and as an NIHR senior investigator. Naaheed Mukadam, Joshua Stott, and Gill Livingston are supported by UCLH National Institute for Health Research (NIHR) Biomedical Research Centre.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY

The data supporting the findings of this study are available within the article (Table 1) and its supplementary material (Supplementary Tables 3–5).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/ADR-230024.

REFERENCES

[1] Gauthier S, Rosa-Neto P, Morais JA, Webster C (2021) *World Alzheimer Report 2021: Journey through the diagnosis of dementia*. Alzheimer's Disease International, London, UK.

[2] ADI (2014) *Dementia in the Asia Pacific Region*, Technical Report, Alzheimer's Disease International, Dementia, Australia.

[3] Statista (2021) The world's most spoken languages.

[4] Shim YS, Yang DW, Kim HJ, Park YH, Kim SY (2017) Characteristic differences in the Mini-Mental State Examination used in Asian countries. *BMC Neurol* **17**, 141.

[5] Xu SH, Chen LH, Jin XQ, Yan J, Xu SZ, Xu Y, Liu CX, Jin Y (2019) Effects of age and education on clock-drawing performance by elderly adults in China. *Clin Neuropsychol* **33**, 96–105.

[6] Huo Z, Lin J, Bat BKK, Chan JYC, Tsoi KKF, Yip BHK (2021) Diagnostic accuracy of dementia screening tools in the Chinese population: A systematic review and meta-analysis of 167 diagnostic studies. *Age Ageing* **50**, 1093-1101.

[7] Wang J, Halffman W, Zwart H (2003) Diagnosing tests: Using and misusing diagnostic and screening tests. *J Pers Assess* **81**, 209-219.

[8] Velayudhan L, Ryu SH, Raczek M, Philpot M, Lindesay J, Critchfield M, Livingston G (2014) Review of brief cognitive tests for patients with suspected dementia. *Int Psychogeriatr* **26**, 1247-1262.

[9] Custodio N, Duque L, Montesinos R, Alva-Diaz C, Mellado M, Slachevsky A (2020) Systematic review of the diagnostic validity of brief cognitive screenings for early dementia detection in Spanish-speaking adults in Latin America. *Front Aging Neurosci* **12**, 270.

[10] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **372**, n71.

[11] McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, Cohen JF, Deeks JJ, Gatsonis C, Hooft L, Hunt HA, Hyde CJ, Korevaar DA, Leeflang MMG, Macaskill P, Reitsma JB, Rodin R, Rutjes AWS, Salameh JP, Stevens A, Takwoingi Y, Tonelli M, Weeks L, Whiting P, Willis BH (2018) Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *JAMA* **319**, 388-396.

[12] Deeks J, Bossuyt P, Leeflang M, Takwoingi Y, Flemyng E (2022) *Cochrane handbook for systematic reviews of diag-*

nostic test accuracy. (Version 2.0), London: The Cochrane Collaboration, 2022.

[13] Heneghan C (2009) EBM resources on the new CEBM website. *BMJ Evid Based Med* **14**, 67.

[14] Mirza N, Panagioti M, Waheed MW, Waheed W (2017) Reporting of the translation and cultural adaptation procedures of the Addenbrooke's Cognitive Examination version III (ACE-III) and its predecessors: A systematic review. *BMC Med Res Methodol* **17**, 141.

[15] Power M, Fell G, Wright M (2013) Principles for high-quality, high-value testing. *BMJ Evid Based Med* **18**, 5-10.

[16] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (2022) Chapter 10: Analysing data and undertaking meta-analyses. In *Cochrane handbook for systematic reviews of interventions version 6.3 (updated February 2022)*.

[17] Rutter CM, Gatsonis CA (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* **20**, 2865-2884.

[18] Nyaga VN, Arbyn M (2022) Metadta: A Stata command for meta-analysis and meta-regression of diagnostic test accuracy data – A tutorial. *Arch Public Health* **80**, 95.

[19] Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* **58**, 882-893.

[20] Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y (2010) Chapter 10: Analysing and presenting results. In *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.*

[21] Chong MS, Lim WS, Chan SP, Feng L, Niti M, Yap P, Yeo D, Ng TP (2010) Diagnostic performance of the Chinese Frontal Assessment Battery in early cognitive impairment in an Asian population. *Dement Geriatr Cogn Disord* **30**, 525-532.

[22] Liew TM, Feng L, Gao Q, Ng TP, Yap P (2015) Diagnostic utility of Montreal cognitive assessment in the fifth edition of diagnostic and statistical manual of mental disorders: Major and mild neurocognitive disorders. *J Am Med Dir Assoc* **16**, 144-148.

[23] Li XD, Jia SH, Zhou Z, Jin Y, Zhang XF, Hou CL, Zheng WJ, Rong P, Jiao JS (2018) The role of the Montreal Cognitive Assessment (MoCA) and its memory tasks for detecting mild cognitive impairment. *Neurol Sci* **39**, 1029-1034.

[24] Malhotra C, Chan A, Matchar D, Seow D, Chuo A, Do YK (2013) Diagnostic performance of short portable mental status questionnaire for screening dementia among patients attending cognitive assessment clinics in Singapore. *Ann Acad Med Singap* **42**, 315-319.

[25] Yeung PY, Wong LL, Chan CC, Leung JL, 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.

[26] Yu RC, Mukadam N, Kapur N, Stott J, Hu CJ, Hong CT, Yang CC, Chan L, Huang LK, Livingston G (2021) Validation of the Taiwanese version of ACE-III (T-ACE-III) to detect dementia in a memory clinic. *Arch Clin Neuropsychol* **37**, 692-703.

[27] Huang L, Chen KL, Lin BY, Tang L, Zhao QH, Lv YR, Guo QH (2018) Chinese version of Montreal Cognitive Assessment Basic for discrimination among different severities of Alzheimer's disease. *Neuropsychiatr Dis Treat* **14**, 2133-2140.

[28] Zhang YR, Ding YL, Chen KL, Liu Y, Wei C, Zhai TT, Wang WJ, Dong WL (2019) The items in the Chinese version of the Montreal cognitive assessment basic discriminate among different severities of Alzheimer's disease. *BMC Neurol* **19**, 269.

[29] Chen CW, Chu H, Tsai CF, Yang HL, Tsai JC, Chung MH, Liao YM, Chi M ju, Chou KR (2015) The reliability, validity, sensitivity, specificity and predictive values of the Chinese version of the Rowland Universal Dementia Assessment Scale. *J Clin Nurs* **24**, 3118-3128.

[30] Fang R, Wang G, Huang Y, Zhuang JP, Tang HD, Wang Y, Deng YL, Xu W, Chen SD, Ren RJ (2014) Validation of the Chinese version of Addenbrooke's cognitive examination-revised for screening mild Alzheimer's disease and mild cognitive impairment. *Dement Geriatr Cogn Disord* **37**, 223-231.

[31] Lee MT, Chang WY, Jang Y (2018) Psychometric and diagnostic properties of the Taiwan version of the Quick Mild Cognitive Impairment screen. *PLoS One* **13**, e0207851.

[32] Li XY, Dai J, Zhao SS, Liu WG, Li HM (2018) Comparison of the value of Mini-Cog and MMSE screening in the rapid identification of Chinese outpatients with mild cognitive impairment. *Medicine* **97**, e10966.

[33] Li X, Shen M, Jin Y, Jia S, Zhou Z, Han Z, Zhang X, Tong X, Jiao J (2021) Validity and reliability of the new Chinese version of the Frontal Assessment Battery-Phonemic. *J Alzheimers Dis* **80**, 371-381.

[34] Low A, Lim L, Lim L, Wong B, Silva E, Ng KP, Kandiah N (2020) Construct validity of the Visual Cognitive Assessment Test (VCAT) - A cross-cultural language-neutral cognitive screening tool. *Int Psychogeriatr* **32**, 141-149.

[35] Shi J, Tian J, Wei M, Miao Y, Wang Y (2012) The utility of the Hopkins Verbal Learning Test (Chinese version) for screening dementia and mild cognitive impairment in a Chinese population. *BMC Neurol* **12**, 1-8.

[36] Wen HB, Zhang ZX, Niu FS, Li L (2008) The application of Montreal cognitive assessment in urban Chinese residents of Beijing. *Zhonghua Nei Ke Za Zhi* **47**, 36-39.

[37] Wong LL, Chan CC, Leung JL, Yung CY, Wu KK, Cheung SYY, Lam CLM (2013) A validation study of the Chinese-Cantonese Addenbrooke's Cognitive Examination Revised (C-ACER). *Neuropsychiatr Dis Treat* **9**, 731-737.

[38] Chang CC, Kramer JH, Lin KN, Chang WN, Wang YL, Huang CW, Lin YT, Chen C, Wang PN (2010) Validating the Chinese version of the Verbal Learning Test for screening Alzheimer's disease. *J Int Neuropsychol Soc* **16**, 244-251.

[39] Chang YT, Chang CC, Lin HS, Huang CW, Chang WN, Lui CC, Lee CC, Lin YT, Chen CH, Chen NC (2012) Montreal cognitive assessment in assessing clinical severity and white matter hyperintensity in Alzheimer's disease with normal control comparison. *Acta Neurol Taiwan* **21**, 64-73.

[40] 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.

[41] Lin KN, Wang PN, Chen C, Chiu YH, Kuo CC, Chuang YY, Liu HC (2003) The three-item clock-drawing test: A simplified screening test for Alzheimer's disease. *Eur Neurol* **49**, 53-58.

[42] Yang LL, Li XJ, Yin J, Yu NW, Liu J, Ye F (2019) A validation study of the Chinese version of the Mini-Addenbrooke's Cognitive Examination for screening mild cognitive impairment and mild dementia. *J Geriatr Psychiatry Neurol* **32**, 205-210.

[43] Xu GL, Meyer JS, Huang YG, Du F, Chowdhury M, Quach M (2003) Adapting Mini-Mental State Examination for dementia screening among illiterate or minimally

educated elderly Chinese. *Int J Geriatr Psychiatry* **18**, 609-616.

[44] Wei M, Shi J, Li T, Ni J, Zhang X, Li Y, Kang S, Ma F, Xie H, Qin B (2018) Diagnostic accuracy of the Chinese version of the trail-making test for screening cognitive impairment. *J Am Geriatr Soc* **66**, 92-99.

[45] Wang B, Ou Z, Gu X, Wei C, Xu J, Shi J (2017) Validation of the Chinese version of Addenbrooke's Cognitive Examination III for diagnosing dementia. *Int J Geriatr Psychiatry* **32**, e173-e179.

[46] Guo Q, Jin L, Hong Z, Lu C (2007) A specific phenomenon of animal fluency test in Chinese elderly. *Chin Mental Health J* **21**, 622-625.

[47] Guo Q, Hong Z, Shi W, Sun Y-M, Lu C (2006) Boston naming test in Chinese elderly, patient with mild cognitive impairment and Alzheimer's dementia. *Chin Mental Health J* **20**, 81-84.

[48] Qian SX, Chen KL, Guan QB, Guo QH (2021) Performance of Mattis dementia rating scale-Chinese version in patients with mild cognitive impairment and Alzheimer's disease. *BMC Neurol* **21**, 1-5.

[49] Guo Q, Hong Z, Yu H, Ding D (2004) Clinical Validity of the Chinese Version of Mattis Dementia Rating Scale in differentiating dementia of Alzheimer type in Shanghai. *Chin J Clin Psychol* **12**, 237-243.

[50] Chen K, Xu Y, Chu A, Ding D, Liang X, Nasreddine ZS, Dong Q, Hong Z, Zhao Q, Guo Q (2016) Validation of the Chinese version of Montreal cognitive assessment basic for screening mild cognitive impairment. *J Am Geriatr Soc* **64**, e285-e290.

[51] Guo QH, Cao XY, Zhou Y, Zhao QH, Ding D, Hong Z (2010) Application study of quick cognitive screening test in identifying mild cognitive impairment. *Neurosci Bull* **26**, 47-54.

[52] Huang L, Chen KL, Lin BY, Tang L, Zhao QH, Li F, Guo QH (2019) An abbreviated version of Silhouettes test: A brief validated mild cognitive impairment screening tool. *Int Psychogeriatr* **31**, 849-856.

[53] Zhao QH, Guo QH, Li F, Zhou Y, Wang B, Hong Z (2013) The Shape Trail Test: Application of a new variant of the Trail making test. *PLoS One* **8**, e57333.

[54] Chan C, Yung C, Pan P (2005) Screening of dementia in Chinese elderly adults by the clock drawing test and the time and change test. *Hong Kong Med J* **11**, 13-19.

[55] Chiu YC, Li CL, Lin KN, Chiu YF, Liu HC (2008) Sensitivity and specificity of the Clock Drawing Test, incorporating Rouleau scoring system, as a screening instrument for questionable and mild dementia: Scale development. *Int J Nurs Stud* **45**, 75-84.

[56] Tsai JC, Chen CW, Chu H, Yang HL, Chung MH, Liao YM, Chou KR (2016) Comparing the sensitivity, specificity, and predictive values of the Montreal Cognitive Assessment and Mini-Mental State Examination when screening people for mild cognitive impairment and dementia in Chinese population. *Arch Psychiatr Nurs* **30**, 486-491.

[57] Tsai PH, Liu JL, Lin KN, Chang CC, Pai MC, Wang WF, Huang JP, Hwang TJ, Wang PN (2018) Development and validation of a dementia screening tool for primary care in Taiwan: Brain Health Test. *PLoS One* **13**, e0196214.

[58] Tsai CF, Lee WJ, Wang SJ, Shia BC, Nasreddine Z, Fuh JL (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. *Int Psychogeriatr* **24**, 651-658.

[59] Tsai JC, Chen CW, Chu H, Yang HL, Chung MH, Liao YM, Chou KR (2016) Comparing the sensitivity, specificity, and predictive values of the Montreal Cognitive Assessment and Mini-Mental State Examination when screening people for mild cognitive impairment and dementia in Chinese population. *Arch Psychiatr Nurs* **30**, 486-491.

[60] Simel DL, Bossuyt PMM (2009) Differences between univariate and bivariate models for summarizing diagnostic accuracy may not be large. *J Clin Epidemiol* **62**, 1292-1300.

[61] Egger M, Higgins JPT, Smith GD (2022) *Systematic reviews in health research: Meta-analysis in context*, John Wiley & Sons Ltd.

[62] Wang J, Halffman W, Zwart H (2021) The Chinese scientific publication system: Specific features, specific challenges. *Learn Publ* **34**, 105-115.