Critical evaluation of COSMIN scores in scales for mild cognitive impairment and Alzheimer's disease: A comprehensive review

Xia Peng[1,#] iD, Run-Chen Li[2,#], Ting Liang[1,#], Shi-Fen Xu[1] iD and Yan Cao[1]

Abstract

Background: Timely diagnosis and intervention of mild cognitive impairment (MCI) can delay the development of Alzheimer's disease (AD).
Objective: The purpose of this study was to analyze assessment tools for cognitive function using the Consensus Criteria for Selection of Health Measurement Instruments (COSMIN) method. Comparing the validity, reliability, and practicality of these assessment tools helps clinicians select appropriate assessment tools for patients, thereby improving diagnostic accuracy.
Methods: We followed the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines and used the COSMIN checklist to conduct a thorough methodological quality assessment of the studies. The measurement properties were evaluated and rated on a scale from excellent to poor, based on adapted criteria. We then synthesized the best evidence by combining the COSMIN outcomes with the quality of findings to ensure a precise and comprehensive analysis.
Results: We identified a total of 156 publications, which included 19 different cognitive assessment instruments. Among these, the Telephone version of the Cantonese Mini-Mental State Examination (T-CMMSE), the Montreal Cognitive Assessment (MoCA), and the Hong Kong versions of the MoCA (HK-MoCA-A1 and A2) demonstrated distinguished qualities. The assessment of measurement properties included internal consistency, reliability, validity, and sensitivity and specificity. Notably, the T-CMMSE showed superior methodological quality based on our rigorous analysis.
Conclusions: The T-CMMSE, MoCA, and HK-MoCA-A1 and A2 were found to be notable cognitive assessment tools for MCI and AD. Future research should aim to expand on these findings by exploring a wider range of tools and contexts.

Introduction

Mild cognitive impairment (MCI) represents a crucial point in the spectrum of cognitive health, signifying a departure from normal aging and potentially foreshadowing the onset of dementias such as Alzheimer's disease (AD). With an estimated one third of those over 65 affected by MCI, the condition poses a considerable public health challenge.[1] Precise and early identification of MCI is essential, as it holds the promise of timely interventions that may decelerate its progression.

The subtle nature of MCI symptoms and its overlap with normal cognitive aging complicate its diagnosis. Clinicians and researchers rely on a range of cognitive assessment

[1]Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China
[2]Shanghai University of Traditional Chinese Medicine, Shanghai, China

[#]These authors contributed equally to this work.

Corresponding authors:
Shi-Fen Xu, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China.
Email: xu_teacher2006@126.com

Yan Cao, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China.
Email: liyangcao@126.com

tools, each with its own strengths and limitations, to distinguish MCI from typical aging. Central to the efficacy of these instruments, like the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), are their psychometric properties, such as sensitivity, specificity, and reliability.

This review aims to conduct a systematic evaluation of these scales, incorporating the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) scoring system—a framework that rates the quality of tools based on measurement properties including internal consistency, structural validity, and reliability.[2] By examining these properties across various MCI scales, the review seeks to determine which instruments provide the most accurate diagnostic capabilities for MCI in different patient populations and care settings. The findings have the potential to significantly impact clinical practice, improving the accuracy of MCI diagnostics and, ultimately, patient care strategies.

Methods

Design

This systematic review adopts the refined COSMIN approach, a methodology tailored for systematic evaluations of Patient-Reported Outcome Measures.[3–5] Our methodology adheres to the updated guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and conforms to AMSTAR (A MeaSurement Tool to Assess systematic Reviews) criteria, which assess the methodological rigor of systematic review.[6,7] This approach ensures a comprehensive, standardized, and transparent review process, allowing for an objective and reliable evaluation of the included studies.

Search strategy

Our search strategy (Figure 1) systematically identified studies relevant to the assessment of MCI using various scales, with a particular focus on the COSMIN scores for these tools. Comprehensive searches were conducted across databases, including PubMed, Web of Science, and Embase, from their inception until December 2023. We used a combination of keywords and MeSH terms tailored to the constructs of interest. Keywords included "Mild Cognitive Impairment," "MCI screening," "COSMIN," "neuropsychological tests," "internal consistency," "reliability," "validity," "sensitivity," and "specificity." Boolean operators such as "AND" and "OR" were used to broaden or narrow the searches as necessary. For example, our search strings included combinations like ("MCI" OR "cognitive impairment, mild") AND "assessment scales" AND "COSMIN scoring" OR "psychometric properties." All identified records underwent a rigorous screening process, and

the search strategy was supplemented by reviewing the references of key articles. The specifics of the search strings, including the full array of search terms and the logic of their combination, are detailed in Figure 1 and Tables 1–7.

Eligibility criteria

A screening process based on titles and abstracts was first conducted to exclude irrelevant studies. Subsequently, articles that potentially met the inclusion criteria were independently evaluated by two reviewers through a comprehensive full-text examination. Conflicts regarding the eligibility of studies were resolved through discussion and consultation with a third reviewer with significant expertise. Inclusion criteria encompassed quantitative studies, including both randomized controlled trials and observational studies, published as full-text original articles. Exclusions were made for brief communications such as letters, editorials, and conference abstracts, focusing the review on comprehensive scholarly reports.

Participants

Patients who received a confirmed diagnosis of MCI or AD by clinical assessment, neuropsychological testing, magnetic resonance imaging or positron emission tomography scans, blood tests, and/or psychiatric evaluation.

Language

The search for studies was confined to those documented in either English or Spanish.

Setting

A manual search was conducted to identify relevant studies beyond electronic databases. The initial screening of studies was jointly performed by two investigators based on the study titles and abstracts. This step was followed by an independent, detailed review of the full texts by a separate pair of researchers. A collaborative approach was used to resolve any disagreements. In cases of multiple reports originating from a single study, characterized by overlapping authorship, methodology, and patient cohorts, priority was given to the initial report. Subsequent reports were only considered if they provided additional insights or outcomes.

Data extraction

Data from included studies were extracted and compiled into a structured template, including variables such as sample size, mean age of participants, gender, study
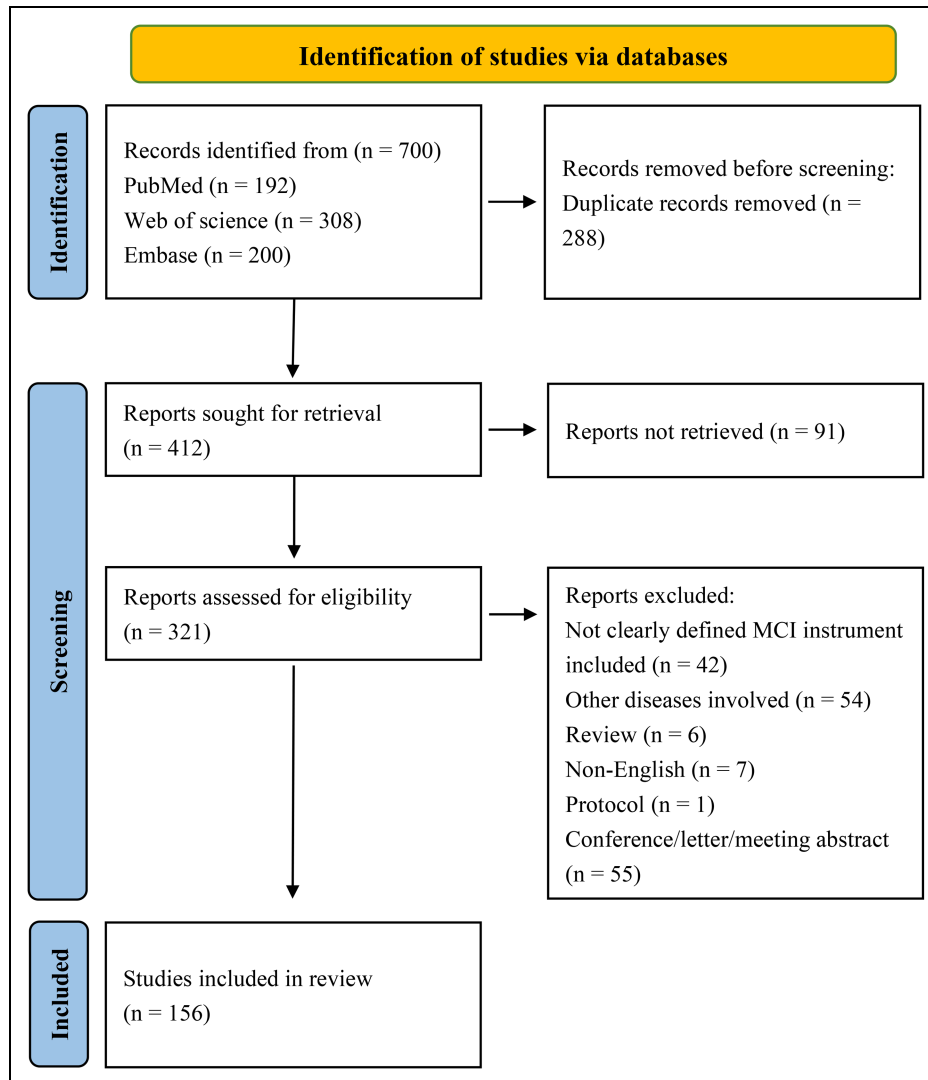
Figure 1. Study flow chart.

setting, country, language, and details of subscales, including the number of items and scale parameters. Additionally, we used the Excel template recommended by COSMIN, available on their official website, to record our evaluations and to assist in determining the methodological quality ratings for each domain of interest.

Handling missing data

In instances where data were incomplete or combined with data from patients with other diseases, we contacted the original authors directly, requesting the specific data related to MCI or AD cohorts. This step ensured the accuracy and comprehensiveness of our dataset for analysis.

Evaluation of the measurement property result studies

The evaluation of the studies' measurement properties was conducted using the COSMIN methodology, which covers a wide range of measurement properties, including but not limited to internal consistency, reliability, measurement error, various aspects of validity (content, construct, cross-cultural, and criterion), responsiveness, and interpretability. Each property within a study was carefully assessed and assigned a quality rating - 'excellent', 'good', 'fair', or 'poor'. The depth of COSMIN analysis varied per study, depending on the extent of measurement properties examined, resulting in different assessments across the reviewed literature.

Evaluation of the methodological quality of the studies

The methodological rigor of the studies was rated using the COSMIN framework, which defines four levels – poor, fair, good, and excellent.[24] This evaluation process was carried out by two independent researchers, with each measurement property being assigned an overall grade based on its lowest individual score. A third expert reviewer was consulted to

Table 1. Measurement properties of performance-based measures: validity.

	Instrument	reference	patients	design	results	correlation	quality	COSMIN
Construct validity	MMSE-2	Baek et al., 2016[8]	226 with MCI	Three factor analysis	0.522 variance	—	+	poor
			97 with AD	Two factor analysis	0.453variance	—	+	
	T-CMMSE	Wong et al., 2009[9]	65	Pearson's Correlation	0.991(p < 0.001)	CMMSE	+	good
	rMMSE-T	Keskinoglu et al., 2009[10]	490	ROC curve area	Educated 0.953 Uneducated 0.907	—	+	good
	A-MMSE	El-Hayeck et al., 2019[11]	782	ROC curve area	0.920	—	+	good
	MoCA-S	Aguilar-Navarro et al., 2018[12]	168	Spearman's correlation	0.830	MMSE	+	poor
Concurrent validity	SMMSE	Mougias et al., 2018[13]	210	Correlation coefficient	Rho 0.820 (p < 0.001)	MMSE	+	poor
					Rho -0.597 (p < 0.001)	GDS	-	
					Rho 0.453 (p < 0.001)	ADL	-	
	HK-MoCA-A1	Wong et al., 2018[14]	60	Pearson's Correlation	0.87(p < 0.001)	HK-MoCA-O	+	good
	HK-MoCA-A2	Wong et al., 2018[14]	60	Pearson's Correlation	0.790(p < 0.001)	HK-MoCA-O	+	good
	MoCA	Gierus et al., 2015[15]	221	Spearman's correlation	Rho 0.926(p = 0.000)	MoCA 7.2 polish version	+	excellent
Convergent validity	S-MoCA-C2	Bezdicek et al., 2020[16]	699	Pearson's Correlation	0.937(p < 0.001)	MoCA-C2	+	excellent

A-MMSE: Arabic-MMSE; AD: Alzheimer's disease; ADL: Activities of Daily Living; CMMSE: Chinese Mini-Mental State Examination; COSMIN: Consensus Criteria for Selection of Health Measurement Instruments; GDS: Global Deterioration Scale; HK-MoCA: Hong Kong version of MoCA; MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; MoCA-S: Spanish Version of the MoCA; rMMSE-T: revised MMSE-Turkish; ROC: receiver operating characteristic; S-MoCA-C2: Short Form Montreal Cognitive Assessment Czech Version; SMMSE: Severe MMSE; T-CMMSE: Telephone version of the Cantonese Mini-Mental State Examination.

mediate and resolve any analytic discrepancies that arose between the initial assessors, thereby ensuring a thorough and objective evaluation of each study's methodological integrity.

Best evidence synthesis: Levels of evidence

Our analysis integrated three distinct dimensions to evaluate the robustness of each measurement property in the assessed questionnaires. First, we employed the COSMIN checklist to determine the methodological quality (detailed in Table 6). Second, we assessed the outcomes of the measurement properties. Finally, we applied the criteria set forth by Terwee et al.[25] (detailed in Tables 1–7). This comprehensive approach allowed for a thorough evaluation of the measurement properties.

In cases where clinical and methodological homogeneity existed among the studies, we employed quantitative synthesis to summarize the outcome measures. Conversely, when substantial heterogeneity was present across the studies, we applied qualitative summary techniques. This dual approach facilitated a nuanced and in-depth understanding of the

measurement properties within the scope of our review, enabling us to account for both the commonalities and the differences among the included studies.

By integrating these three dimensions—methodological quality, assessment of measurement property outcomes, and the application of standardized criteria—our analysis provided a robust and comprehensive evaluation of the questionnaires. This methodology not only allowed for a systematic appraisal of the measurement properties but also ensured that the findings were grounded in a well-established framework, enhancing the reliability and validity of our conclusions.

Results

Measurement properties

Internal consistency. The internal consistency of various MCI assessment scales was examined using the COSMIN checklist to evaluate their methodological quality and reliability of each instrument was assessed with Cronbach's alpha, ensuring consistent measurement of cognitive constructs.

Table 2. Measurement properties of performance-based measures: reliability.

Instrument	Reference	Patients	Interval	Design	Measurement	Results	Quality	COSMIN
MMSE-2	Baek et al., 2016[8]	323	34.48 ± 3.48	Test-retest analysis	ICC	0.76–0.90	+	poor
SMMSE	Mougias et al., 2018[13]	20	17.20 ± 5.17	Test-retest analysis	Spearman's correlation	Rho 0.944 p < 0.001	+	poor
T-CMMSE	Wong et al., 2009[9]	65	7	Test-retest analysis	ICC	0.99 p < 0.001	+	excellent
rMMSE-T	Keskinoglu et al., 2009[10]	490	-	Test-retest analysis	ICC	0.966 p < 0.001 educated	+	good
				Test-retest analysis	ICC	0.988 p < 0.001 uneducated	+	
A-MMSE	El-Hayeck et al., 2019[11]	782	-	Test-retest analysis	ICC	0.89 p < 0.001	+	poor
H-MoCA	Gupta et al., 2019[17]	30	30	Test-retest analysis	ICC	0.87	+	good
MoCA-S	Aguilar-Navarro et al., 2018[12]	168	-	Test-retest analysis	ICC	0.86	+	poor
T-MoCA	Karim et al., 2022[18]	233	35	Test-retest analysis	ICC	0.918 p = 0.001	+	poor
Arabic MoCA	Rahman et al., 2009[19]	184	35 ± 17.6	Test-retest analysis	Score	0.9–2.5	+	poor
HK-MoCA-A1	Wong et al., 2018[14]	60	30	Test-retest analysis	ICC	0.99 p < 0.001	+	good
HK-MoCA-A2	Wong et al., 2018[14]	60	30	Test-retest analysis	ICC	0.82 p < 0.001	+	good

A-MMSE: Arabic-MMSE; COSMIN: Consensus Criteria for Selection of Health Measurement Instruments; H-MoCA: Hindi translated version of the MoCA; HK-MoCA: Hong Kong version of MoCA; ICC: intraclass correlation coefficient; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; MoCA-S: Spanish Version of the MoCA; rMMSE-T: revised MMSE-Turkish; SMMSE: Severe MMSE; T-CMMSE: Telephone version of the Cantonese Mini-Mental State Examination.; T-MoCA: Tamil (India) version of MoCA.

Table 3. Measurement properties of performance-based measures: internal consistency.

Instrument	Reference	Patients	Measurement	Results	Quality	COSMIN
MMSE-2	Baek et al., 2016[8]	323	Cronbach's α	0.62–0.79	+	poor
SMMSE	Mougias et al., 2018[13]	210	Cronbach's α	0.87	+	good
rMMSE-T	Keskinoglu et al., 2009[10]	490	Cronbach's α	0.713 (educated)	+	poor
A-MMSE	El-Hayeck et al., 2019[11]	782	Cronbach's α	0.734 (uneducated)		poor
				0.71	+	poor
H-MoCA	Gupta et al., 2019[17]	30	Cronbach's α	0.642	+	poor
MoCA-S	Aguilar-Navarro et al., 2018[12]	168	Cronbach's α	0.891	+	good
T-MoCA	Karim et al., 2022[18]	233	Cronbach's α	0.834	+	poor
Arabic MoCA	Rahman et al., 2009[19]	184	Cronbach's α	0.83	+	poor
HK-MoCA-A1	Wong et al., 2018[14]	60	Cronbach's α	0.79	+	good
HK-MoCA-A2	Wong et al., 2018[14]	60	Cronbach's α	0.75	+	good
Georgian MoCA	Janelidze et al., 2017[20]	86	Cronbach's α	0.86–0.92	+	good
MoCA	Gupta et al., 2019[17]	360	Cronbach's α	0.903	+	good
EIACE	Sarasola et al., 2005[21]	127	Cronbach's α	0.8137	+	poor
P-GPCOG	Rashedi et al., 2019[22]	151	Cronbach's α	0.9 (Community dwelling elders)	+	good
		79		0.83 (Nursing home residents)	+	poor

A-MMSE: Arabic-MMSE; COSMIN: Consensus Criteria for Selection of Health Measurement Instruments; EIACE: El Addenbrooke's Cognitive Examination; H-MoCA: Hindi translated version of the MoCA; HK-MoCA: Hong Kong version of MoCA; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; MoCA-S: Spanish Version of the MoCA; P-GPCOG: Persian version of general practitioner assessment of cognition; rMMSE-T: revised MMSE-Turkish; SMMSE: Severe MMSE; T-MoCA: Tamil (India) version of MoCA.

Several cognitive assessment tools demonstrated satisfactory to excellent internal consistency, as evidenced by their Cronbach's alpha values and COSMIN scores. The MMSE-2, which involved 323 participants, demonstrated a Cronbach's alpha between 0.62 and 0.79.[8] This range indicates satisfactory internal consistency, as evidenced

Table 4. Measurement properties of performance-based measures: sensitivity and specificity.

Instrument	Reference	Patients	Measurement	Best cut off score	Sensitivity for MCI	Specificity for MCI	Sensitivity for AD	Specificity for AD	Quality	COSMIN
T-CMMSE	Wong et al., 2009[9]	233	ROC	16 of 26	100%	96.7%	—	—	+	excellent
MMSE-2SV	Baek et al., 2016[8]	323	ROC	26 of 30	74%	59%	—	—	-	poor
		323	ROC	23 of 30			92%	87%	-	good
rMMSE-T	Keskinoglu et al., 2009[10]	490	ROC	22 (educated)	90.9%	97%	—	—	+	good
			ROC	19 (uneducated)	82.7%	92.3%	—	—	+	fair
A-MMSE	El-Hayeck et al., 2019[11]	782	ROC	23 of 30	80%	89.4%	—	—	+	poor
MoCA-S	Aguilar-Navarro et al., 2018[12]	168	ROC	26 of 30	80%	75%	—	—	+	poor
Arabic MoCA	Rahman et al., 2009[19]	184	ROC	26 of 30	92.3%	85.7%	—	—	+	poor
T-MoCA	Karim et al., 2022[18]	233	ROC	24 of 30	88.4%	77.9%	—	—	+	poor
Georgian MoCA	Janelidze et al., 2017[20]	86	ROC	22 of 30	100%	69%	—	—	+	excellent
MoCA	Gierus et al., 2015[15]	221	ROC	23 of 30	82.6%	69.7%	—	—	+	poor
MoCA	Gupta et al., 2019[17]	360	ROC	22 of 30	81%	77%			+	poor
		360	ROC	17 of 30	—	—	88%	98%	+	excellent
EIACE	Sarasola et al., 2005[21]	128	ROC	86 of 100	92.1%	96.2%	—	—	+	good
ACE-III	Potts et al., 2022[31]	2176	ROC	84 of 100	92%	63%	—	—	±	fair
Qmci-J	Morita et al., 2019[23]	526	ROC	60 of 100	73%	68%	—	—	±	poor

A-MMSE: Arabic-MMSE; ACE-III: Addenbrooke's Cognitive Examination-III; AD: Alzheimer's disease; COSMIN: Consensus Criteria for Selection of Health Measurement Instruments; EIACE: EI Addenbrooke's Cognitive Examination; MCI: mild cognitive impairment; MMSE-2SV: The MMSE-2; MMSE-2: Standard Version; MoCA: Montreal Cognitive Assessment; MoCA-S: Spanish Version of the MoCA; Qmci-J: Japanese version of the Quick Mild Cognitive Impairment; rMMSE-T: revised MMSE-Turkish; ROC: receiver operating characteristic; T-MoCA: Tamil (India) version of MoCA; T-CMMSE: Telephone version of the Cantonese Mini-Mental State Examination.

Table 5. Summary of measurement properties.

Instrument	Validity	Reliability	Internal consistency	Sensitivity and Specificity
MMSE-2	?	?	?	?
T-CMMSE	++	+++	/	+++
rMMSE-T	++	++	?	+
A-MMSE	++	?	?	?
SMMSE	?	?	++	/
MoCA-S	?	?	++	?
HK-MoCA-A1	++	++	++	/
HK-MoCA-A2	++	++	++	/
MoCA	+++	/	++	+/-
S-MoCA-C2	+++	/	/	/
H-MoCA	/	++	?	/
T-MoCA	/	?	?	?
Arabic MoCA	/	?	?	?
Georgian MoCA	/	/	++	+++
EIACE	/	/	?	++
ACE-III	/	/	/	+
Qmci-J	/	/	/	?
P-GPCOG	/	/	?	/

+++, consistent findings of positive rating in multiple studies of 'good' methodological quality OR in one study of 'excellent' methodological quality; + +, consistent findings of positive rating in multiple studies of 'fair' methodological quality OR in one study of 'good' methodological quality; +, positive result in one study of 'fair' methodological quality; +/-, conflicting findings; ?, only studies of 'poor' methodological quality; – – –, consistent findings of negative rating in multiple studies of 'good' methodological quality OR in one study of 'excellent' methodological quality; – –, consistent findings of negative rating in multiple studies of 'fair' methodological quality OR in one study of 'good' methodological quality; –, negative result in one study of 'fair' methodological quality; O, no evidence.

A-MMSE: Arabic-MMSE; ACE-III: Addenbrooke's Cognitive Examination-III; EIACE: El Addenbrooke's Cognitive Examination; H-MoCA: Hindi translated version of the MoCA; HK-MoCA: Hong Kong version of MoCA; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; MoCA-S: Spanish Version of the MoCA; P-GPCOG: Persian version of general practitioner assessment of cognition; Qmci-J: Japanese version of the Quick Mild Cognitive Impairment; rMMSE-T: revised MMSE-Turkish; S-MoCA-C2: Short Form Montreal Cognitive Assessment Czech Version; SMMSE: Severe MMSE; T-CMMSE: Telephone version of the Cantonese Mini-Mental State Examination.; T-MoCA: Tamil (India) version of MoCA.

by a positive COSMIN score. Similarly, the Severe MMSE (SMMSE) sampled 210 individuals and reported a Cronbach's alpha of 0.87, suggesting high internal consistency, supported by a positive COSMIN score.[13]

The revised MMSE-Turkish (rMMSE-T), administered to 490 patients, yielded Cronbach's alpha values of 0.713 for educated subjects and 0.734 for uneducated ones, both with positive COSMIN scores.[10] The Arabic-MMSE (A-MMSE), applied to 782 patients, yielded a Cronbach's alpha of 0.71, confirming consistent performance with a positive COSMIN rating.[11] The Hindi translated version of the MoCA (H-MoCA), administered to 30 individuals, resulted in a Cronbach's alpha of 0.642, indicating moderate internal consistency but a poor COSMIN score, suggesting the need for further validation.[17] In contrast, the Spanish Version of the MoCA (MoCA-S), involving 168 subjects, demonstrated high internal consistency (Cronbach's alpha = 0.891) but a poor COSMIN score.[12] The Tamil (India) version of MoCA (T-MoCA) reported a Cronbach's alpha of 0.834 from 233 participants, along with a positive COSMIN score.[18]

Furthermore, the Arabic MoCA and the Hong Kong version of MoCA (HK-MoCA-A1) demonstrated good internal consistency. The Arabic MoCA (184 participants) and HK-MoCA-A1 (60 individuals) exhibited good internal consistency, with Cronbach's alpha values of 0.83 and 0.79, respectively, and positive COSMIN scores.[14,19] The HK-MoCA-A2 (60 patients) and Georgian MoCA (86 participants) showed Cronbach's alpha values of 0.75 and 0.86–0.92, respectively.[20] The MoCA, with a larger sample of 360 patients, displayed excellent internal consistency (Cronbach's alpha = 0.903) and a positive COSMIN score.[26] Finally, the El Addenbrooke's Cognitive Examination (EIACE; 127 subjects) and Persian version of general practitioner assessment of cognition (P-GPCOG; assessed in 151 community-dwelling elders and 79 nursing home residents) demonstrated high internal consistency, with Cronbach's alpha values of 0.8137, 0.9, and 0.83, respectively, and positive COSMIN ratings.[21,22]

Reliability

Several cognitive assessment tools were evaluated for their test-retest reliability to ensure consistency of results over time.

The MMSE-2, involving 323 participants, showed an impressive intraclass correlation coefficient (ICC) between 0.76 and 0.90, indicating high reliability. The SMMSE,

Table 6. Articles included in the review with COSMIN methodological evaluation scores.

Instrument	References	Properties	COSMIN
MMSE-2	Baek et al., 2016[8]	Construct validity	poor
		Reliability	poor
		Internal consistency	poor
		Sensitivity and Specificity	poor
T-CMMSE	Wong et al., 2009[9]	Construct validity	good
		Reliability	excellent
		Sensitivity and Specificity	excellent
rMMSE-T	Keskinoglu et al., 2009[10]	Construct validity	good
		Reliability	good
		Internal consistency	poor
		Sensitivity and Specificity	fair
A-MMSE	El-Hayeck et al., 2019[11]	Construct validity	good
		Reliability	poor
		Internal consistency	poor
		Sensitivity and Specificity	poor
SMMSE	Mougias et al., 2018[13]	Concurrent validity	poor
		Reliability	poor
		Internal consistency	good
MoCA-S	Aguilar-Navarro et al., 2018[12]	Construct validity	poor
		Reliability	poor
		Internal consistency	good
		Sensitivity and Specificity	poor
HK-MoCA-A1	Wong et al., 2018[14]	Concurrent validity	good
		Reliability	good
		Internal consistency	good
HK-MoCA-A2	Wong et al., 2018[14]	Concurrent validity	good
		Reliability	good
		Internal consistency	good
MoCA	Gierus et al., 2015[15]	Concurrent validity	excellent
		Sensitivity and Specificity	poor
	Gupta et al., 2019[17]	Internal consistency	good
		Sensitivity and Specificity	poor
S-MoCA-C2	Bezdicek et al., 2020[16]	Convergent validity	excellent
H-MoCA	Gupta et al., 2019[17]	Reliability	good
		Internal consistency	poor
T-MoCA	Karim et al., 2022[18]	Reliability	poor
		Internal consistency	poor
		Sensitivity and Specificity	poor
Arabic MoCA	Rahman et al., 2009[19]	Reliability	poor
		Internal consistency	poor
		Sensitivity and Specificity	poor
Georgian MoCA	Janelidze et al., 2017[20]	Internal consistency	good
		Sensitivity and Specificity	excellent
EIACE	Sarasola et al., 2005[21]	Internal consistency	poor
		Sensitivity and Specificity	good
ACE-III	Potts et al., 2022[31]	Sensitivity and Specificity	fair
Qmci-J	Morita et al., 2019[23]	Sensitivity and Specificity	poor
P-GPCOG	Rashedi et al., 2019[22]	Internal consistency	poor

A-MMSE: Arabic-MMSE; ACE-III: Addenbrooke's Cognitive Examination-III; EIACE: EI Addenbrooke's Cognitive Examination; H-MoCA: Hindi translated version of the MoCA; HK-MoCA: Hong Kong version of MoCA; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; MoCA-S: Spanish Version of the MoCA; P-GPCOG: Persian version of general practitioner assessment of cognition; Qmci-J: Japanese version of the Quick Mild Cognitive Impairment; rMMSE-T: revised MMSE-Turkish; S-MoCA-C2: Short Form Montreal Cognitive Assessment Czech Version; SMMSE: Severe MMSE; T-CMMSE: Telephone version of the Cantonese Mini-Mental State Examination.; T-MoCA: Tamil (India) version of MoCA.

with 20 subjects, demonstrated a strong Spearman's correlation coefficient (rho) of 0.944 ($p < 0.001$), reflecting consistent performance in repeated assessments.

The Telephone version of the Cantonese Mini-Mental State Examination (T-CMMSE), including 65 participants, achieved an outstanding ICC of 0.99 ($p < 0.001$),

Table 7. Details of the included instruments.

Instrument	Orientation	Attention	Computation	Recall	immediate memory	Visuospatial function	language	Drawing clock test	Others
MMSE	+	+	+	+	+	-	+	-	—
MoCA	+	+	+	+	-	+	+	-	Executive ability, word abstract generalization ability
Qmci	+	-	-	+	+	-	+	+	logic
ACE-III	+	+	-	+	-	+	+	-	Intonation fluency
P-GPCOG	+	-	-	+	-	-	-	+	Name address memory, information

ACE-III: Addenbrooke's Cognitive Examination-III; P-GPCOG, Persian version of General Practitioner Assessment of Cognition; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; Qmci: Quick Mild Cognitive Impairment.

indicating excellent test-retest reliability and supported by an 'excellent' COSMIN score.[9] This finding was further supported by an 'excellent' COSMIN score. The rMMSE-T, assessing 490 individuals, revealed an ICC of 0.966 (p < 0.001) for educated participants and 0.988 (p < 0.001) for uneducated participants, both indicating very high reliability.

The A-MMSE was evaluated with 782 patients and reported an ICC of 0.89 (p < 0.001), ensuring its reliability in a large and diverse sample. The H-MoCA, involving 30 subjects, showed an lCC of 0.87 (p < 0.001), which is considered 'good' according to COSMIN criteria.

The MoCA-S, which included 168 participants, had an lCC of 0.86. The T-MoCA, with 233 subjects, reported an ICC of 0.918 (p = 0.001), both suggesting strong reliability. The Arabic MoCA, assessed with 184 individuals, utilized a scoring system and showed results ranging from 0.9 to 2.5, affirming its reliability over time.

The HK-MoCA-A1, with 60 participants, demonstrated an excellent ICC of 0.99 (p < 0.001), rated 'good' by COSMIN standards. The subsequent version, HK-MoCA-A2, with the same number of participants, also reported a 'good' ICC of 0.82 (p < 0.001).

These assessments collectively confirm the robustness and reliability of the scales for assessing cognitive impairment. Most instruments exhibited high internal consistency and received positive COSMIN ratings, underscoring their suitability for repeated use in clinical and research settings.

Validity

Various cognitive assessment tools for MCI and AD were evaluated for their construct, concurrent, and convergent validity. The MMSE-2, involving 323 participants, demonstrated satisfactory construct validity with variances of 0.522 and 0.453 in three-factor and two-factor analyses, respectively. These results, set against measures of cognitive function, indicate satisfactory construct validity. The T-CMMSE, assessed among 65 individuals, showed an impressive Pearson's correlation of 0.991 (p < 0.001) with the standard MMSE, suggesting excellent construct validity. Similarly, the rMMSE-T, administered to 490 individuals, exhibited excellent construct validity with receiver operating characteristic (ROC) curve areas of 0.953 for the educated group and 0.907 for the uneducated group. The A-MMSE was administered to 782 patients and showed a robust ROC curve area, indicative of strong construct validity.

Concurrent validity was examined using the SMMSE with 210 patients, demonstrating a Spearman's correlation coefficient of 0.820 (p < 0.001) against the MMSE.[27] It also showed correlations of −0.597 (p < 0.001) with the Global Deterioration Scale (GDS) and 0.453 (p < 0.001) with the Activities of Daily Living (ADL) scale, confirming its effectiveness in reflecting concurrent cognitive and functional statuses. HK-MoCA-A1 and HK-MoCA-A2, with 60 subjects each, yielded Pearson's correlations of 0.87 and 0.79 (p < 0.001) against the HK-MoCA-O, respectively, substantiating their concurrent validity. The MoCA, with 221 participants, showed a strong Spearman's correlation of 0.926 (p < 0.001) with the MoCA 7.2 Polish version, highlighting its concurrent validity. Moreover, the MoCA-S, involving 168 subjects, exhibited a Spearman's correlation of 0.830 with the MMSE, further supporting its concurrent validity.

Convergent validity was assessed with the Short Form MoCA Czech Version (S-MoCA-C2), involving 699 individuals, revealing a Pearson's correlation of 0.937 (p < 0.001) with the MoCA-C2, indicating high convergent validity.

Sensitivity and specificity

A comprehensive evaluative review of MCI assessment scales examined a range of instruments for their diagnostic

sensitivity and specificity. The T-CMMSE with 233 patients showed a sensitivity and specificity for MCI of 100% and 96.7%, respectively, at the optimal cut-off score of 16 out of 26. The MMSE-2: Standard Version (MMSE-2SV), involving 323 individuals, demonstrated a sensitivity of 74% for MCI and 92% for AD, with specificity at 59% for MCI and 87% for AD, at a cut-off score of 26 out of 30.

The rMMSE-T, administered to 490 subjects, reported sensitivities of 90.9% for educated and 82.7% for uneducated individuals, and specificities of 97% and 92.3% for MCI, respectively, at the cut-offs of 22 and 19. The A-MMSE, with a large cohort of 782 patients, showed 80% sensitivity and 89.4% specificity for MCI at a cut-off of 23 out of 30. The MoCA-S, assessed among 168 participants, exhibited 80% sensitivity and 75% specificity for MCI at a cut-off score of 26 out of 30. The Arabic version of the MoCA, with 184 subjects, had a sensitivity of 92.3% and specificity of 85.7% for MCI at the cut-off of 26 out of 30.

The T-MoCA, involving 233 participants, presented an 88.4% sensitivity and 77.9% specificity for MCI at a cut-off of 24 out of 30. The Georgian MoCA, with 86 individuals, reported 100% sensitivity and 69% specificity for MCI at a cut-off of 22 out of 30.

Two separate studies evaluated the standard MoCA. The first, involving 221 patients, showed a sensitivity of 82.6% and specificity of 69.7% for MCI at a cut-off of 23 out of 30, while the second, with 360 participants, demonstrated 81% sensitivity and 77% specificity for MCI at a cut-off of 22 out of 30. For AD, the second study displayed remarkable sensitivity (88%) and specificity (98%) at a cut-off of 17 out of 30.

Additionally, the EIACE, assessed with 128 subjects, revealed a sensitivity of 92.1% and specificity of 96.2% for MCI at a cut-off of 86 out of 100. The Addenbrooke's Cognitive Examination-III (ACE-III), with a substantial sample of 2176 individuals, showed a sensitivity of 92% and specificity of 63% for MCI at a cut-off of 84 out of 100. The Japanese version of the Quick Mild Cognitive Impairment (Qmci-J), evaluated among 526 subjects, reported a sensitivity of 73% and specificity of 68% for MCI at a cut-off of 60 out of 100.[23]

The COSMIN ratings varied across instruments, with several scales rated as 'excellent' or 'good', signifying robust methodological quality. These findings underscore the importance of selecting scales with proven sensitivity and specificity for both MCI and AD, ensuring precise and reliable cognitive impairment assessments in clinical practice.

Methodological evaluation of measurement properties

The COSMIN criteria were used to evaluate the methodological quality of cognitive assessment tools, revealing a spectrum of methodological robustness. The MMSE-2, T-CMMSE, and SMMSE displayed methodological shortcomings, receiving predominantly 'poor' ratings. In contrast, the A-MMSE demonstrated more favorable outcomes, with ratings ranging from 'good' to 'fair.' The MoCA-S, Georgian MoCA, HK-MoCA-A1 and A2 exhibited 'good' ratings across several measurement properties, indicating their sound methodological construct. The rMMSE-T and S-MoCA-C2 were distinguished by their 'good' methodological quality. Variability was observed in the H-MoCA, T-MoCA, and Arabic MoCA, which presented a dichotomy of 'good' and 'poor' ratings, suggesting the need for cautious application. The standard MoCA, through separate validations, consistently showed 'good' to 'excellent' ratings. Conversely, the EIACE, P-GPCOG, and Qmci-J exhibited a juxtaposition of 'good' ratings in some respects, yet 'poor' in others. The ACE in Spanish and ACE-III, while not uniformly robust, showed promise in certain domains with 'good' ratings, underscoring their selective utility. This comprehensive analysis underscores the imperative for judicious selection of cognitive assessment instruments, ensuring alignment with clinical and research imperatives for accurate cognitive evaluation.

Best evidence synthesis: Overall levels of evidence

The methodological caliber of cognitive assessment instruments, appraised via COSMIN standards, exhibits a spectrum of efficacy. The MMSE-2, T-CMMSE, A-MMSE, SMMSE, and MoCA-S are characterized by a range of methodological qualities, varying from poor to fair, indicating the need for cautious interpretation of their results. In contrast, the rMMSE-T, HK-MoCA-A1 and A2 are noted for their reliability, with consistently good ratings across multiple measurement properties, making them more suitable for repeated assessments.

The standard MoCA stands out with its 'excellent' validity ratings, underscoring its methodological integrity and suitability for accurate cognitive evaluations. Similarly, the Georgian MoCA is distinguished by its superior sensitivity and specificity, making it a valuable tool for identifying cognitive impairment in clinical settings.

However, the Arabic MoCA, EIACE, ACE in Spanish, ACE-III, Qmci-J, and P-GPCOG necessitate judicious use due to their varied methodological ratings, ranging from poor to good across different measurement properties. This variability highlights the importance of carefully considering the specific strengths and limitations of each tool before application.

The S-MoCA-C2's 'excellent' validity enhances its methodological profile, suggesting its potential for accurate cognitive assessments. In contrast, the H-MoCA and T-MoCA results remain indeterminate, with a mix of good and poor ratings, indicating the need for further validation studies to establish their methodological robustness.

This synthesis accentuates the importance of discerning the appropriate cognitive assessment tool, contingent upon

the COSMIN ratings, to ensure precise evaluation of cognitive impairment. By carefully considering the methodological quality of each instrument, clinicians and researchers can select the most appropriate tool for their specific needs, optimizing the accuracy and reliability of cognitive assessments in diverse populations.

Discussion

Studies have revealed significant variations in the prevalence of MCI and its progression to AD. For instance, a study conducted on subjects with MCI revealed that 53% had prodromal AD according to International Working Group-1 criteria. The three-year progression rate to AD-type dementia in this group was 50%, compared to 21% for subjects without prodromal AD.[28] This information highlights the importance of accurate diagnosis and timely intervention. Additionally, the prevalence of MCI varies based on demographic and geographical factors. A study in elderly Chinese populations found the overall prevalence of MCI to be 20.8%, with a higher incidence in rural (23.4%) compared to urban areas (16.8%),[29] suggesting that factors such as lifestyle and access to healthcare could play a role in the incidence and identification of MCI. These findings highlight the need for cognitive assessment tools that are sensitive and adaptable to different populations and stages of cognitive decline, as well as the necessity for tools that can accurately differentiate between MCI and early stages of AD.

In the area of neurocognitive disorders, particularly MCI and AD, precise cognitive assessment is the cornerstone of diagnosis. The complex clinical presentation of MCI, which often overlaps with the spectrum of normal cognitive aging, necessitates tools that are both sensitive and specific. This section explores the pivotal role of standardized cognitive assessment instruments, such as the MMSE-2, T-CMMSE, and MoCA. These tools are examined for their psychometric robustness, including their sensitivity, specificity, and overall reliability.[30] The variation in methodological quality of these instruments, as evidenced by the COSMIN scoring system, underscores the imperativeness of selecting the most appropriate tool tailored to the clinical context. An in-depth discussion of these tools' psychometric properties and their application in diagnosing MCI and AD not only illuminate the complexities inherent in such assessments but also address the need for continuous refinement of these diagnostic instruments. This analysis is crucial in underscoring the role of accurate cognitive assessment as a linchpin in the timely and accurate diagnosis of MCI and AD, ultimately paving the way for targeted therapeutic interventions.

In the complex arena of cognitive assessment for MCI and AD, the HK-MoCA-A1 and HK-MoCA-A2 have emerged as significant instruments. Exhibiting high internal consistency (Cronbach's alpha: HK-MoCA-A1 – 0.79,

HK-MoCA-A2 – 0.75) alongside commendable test-retest reliability, these tools are validated as robust instruments for cognitive screening within the Hong Kong demographic. When compared to other prevalent instruments, the MMSE-2, despite its widespread application, presents limitations in its sensitivity, notably in the early detection of cognitive decline. However, its straightforward administration protocol makes it an optimal instrument for general cognitive screening across a spectrum of settings. In contrast, the T-CMMSE distinguishes itself with its superior reliability, especially in telephonic assessments, thus offering a pragmatic solution for patient populations that are remote or otherwise inaccessible. Furthermore, the MoCA is characterized by its enhanced sensitivity in identifying MCI, rendering it a preferred tool for early-stage diagnosis. Its various iterations, inclusive of the HK-MoCA-A1 and A2, are tailored to address diverse cultural and linguistic nuances, thereby broadening its international applicability. In considering the appropriateness of these tools for distinct demographic groups, the HK-MoCA-A1 and A2 are particularly apt for Cantonese-speaking cohorts, ensuring a culturally and linguistically congruent assessment. The MMSE-2, while less sensitive but user-friendly, is apt for initial screenings spanning a wide range of populations, from community settings to primary care environments. The telephonic adaptation of the T-CMMSE avails itself as a unique asset for populations encumbered by limited access to in-person healthcare services. In conclusion, the selection of a cognitive assessment instrument should be meticulously informed by the specificities of the patient demographic, including but not limited to linguistic, cultural, and healthcare access factors. The findings of this study advocate for a bespoke approach in instrument selection, aiming to optimize the efficacy and accuracy of cognitive assessments for MCI and AD. This revised section provides a more scientifically nuanced and polished overview of the various cognitive assessment tools, emphasizing their suitability for different patient groups and diagnostic contexts.

In the critical appraisal of cognitive assessment tools for MCI and AD, the COSMIN criteria serve as a pivotal framework for methodological evaluation. This framework stratifies the tools into four methodological echelons: poor, fair, good, and excellent, thereby offering a comprehensive perspective on their psychometric properties. For example, the MMSE-2, a widely employed instrument, manifests considerable methodological limitations, as evidenced by its predominant allocation in the 'poor' category across diverse measurement dimensions such as construct validity, reliability, internal consistency, and sensitivity and specificity. In contrast, the T-CMMSE garners 'good' to 'excellent' methodological accolades, particularly notable in the realms of reliability and construct validity, signifying its exemplary methodological integrity. Further, the HK-MoCA-A1 and A2 exhibit commendable methodological merit,

predominantly achieving 'good' ratings in multiple measurement properties. Notably, the HK-MoCA-A1 demonstrates a remarkable ICC of 0.99 in test-retest reliability analyses. The standard MoCA, through its independent validations, consistently secures 'good' to 'excellent' methodological standings, affirming its robustness and reliability. Conversely, instruments such as the EIACE, ACE-III, and P-GPCOG showcase a dichotomy in their methodological ratings, oscillating between 'good' and 'poor'. This dichotomous trend underlines the necessity for cautious application and interpretation in varied clinical and research contexts. Thus, the application of the COSMIN framework in this study underscores the critical need for discerning selection of cognitive assessment instruments. This study emphasizes the critical importance of methodological rigor in ensuring accurate cognitive evaluation, particularly in the context of MCI and AD. The evident variability in methodological quality, as demonstrated through the systematic analysis of various cognitive assessment tools, underscores the necessity for ongoing evaluation and refinement of these instruments. By continually assessing and improving the psychometric properties and methodological robustness of these tools, researchers and clinicians can enhance their applicability and effectiveness across a wide range of patient populations and clinical settings.

Furthermore, the evaluation of cognitive assessment tools for MCI and AD revealed a diverse range of methodological robustness across the assessed instruments. The MMSE-2 and SMMSE predominantly received 'poor' ratings, indicating significant methodological limitations. On the contrary, the T-CMMSE and the rMMSE-T demonstrated superior performance, with the T-CMMSE achieving an outstanding ICC of 0.99, indicative of excellent test-retest reliability, and the rMMSE-T revealing high reliability with ICCs of 0.966 and 0.988 for educated and uneducated participants respectively. In contrast, the Automated MMSE showed more moderate outcomes, with ratings ranging from 'good' to 'fair,' suggesting its relatively lesser methodological robustness compared to the T-CMMSE and rMMSE-T. The MoCA and its adaptations, such as the Georgian MoCA, HK-MoCA-A1 and A2, were distinguished by 'good' to 'excellent' ratings, reflecting their robust methodological constructs. The MoCA, in particular, was noted for its 'excellent' validity, highlighting its methodological integrity and appropriateness for cognitive evaluation. The Georgian MoCA was identified for its outstanding sensitivity and specificity, rendering it a preferred instrument for precise cognitive impairment assessment. However, tools including the Arabic MoCA, EIACE, ACE-III, Qmci-J, and P-GPCOG demonstrated mixed methodological ratings, necessitating careful consideration in their application. The S-MoCA-C2, with its 'excellent' validity, enhances its profile in methodological terms, while the results for the H-MoCA and T-MoCA remain inconclusive, suggesting the need for cautious use.

Based on these findings, the following recommendations are proposed:

1. Initial Screening: For initial cognitive screenings, the MMSE-2 is recommended due to its user-friendly administration, despite its lower sensitivity in detecting early-stage cognitive decline. Its straightforward protocol makes it an ideal choice for primary care settings and community-based assessments.
2. Comprehensive Cognitive Evaluations: When a more detailed cognitive assessment is required, the MoCA and its variants, particularly the Georgian MoCA, are advised. These tools demonstrate superior methodological qualities, including heightened sensitivity and specificity, making them well-suited for in-depth evaluations of cognitive function.
3. Culturally and Linguistically Diverse Populations: For patient populations with specific cultural and linguistic needs, such as Cantonese-speaking individuals, the HK-MoCA-A1 and A2 are recommended. These adaptations ensure that cognitive assessments are conducted in a manner that is compatible with the patient's background, enhancing the accuracy and relevance of the results.
4. Remote or Underserved Populations: In cases where patients have limited access to in-person healthcare services, the T-CMMSE is recommended. This tool has demonstrated strong performance in telephonic assessments, making it an ideal choice for remote cognitive evaluations or for reaching underserved populations.
5. Research Settings: For research applications, instruments that have received 'good' to 'excellent' COSMIN ratings are preferred. Tools such as the MoCA, Georgian MoCA, and S-MoCA-C2 have shown robust methodological quality, ensuring that study findings are reliable and valid. Researchers should prioritize the use of these instruments to maintain high standards of scientific rigor.

In conducting this systematic review, several methodological limitations need to be addressed, highlighting the importance of careful interpretation of the results. A significant constraint is the linguistic scope of the review, limited to literature in English and Spanish. This exclusion of studies in other languages may introduce selection bias, potentially omitting crucial research and affecting the breadth of the analysis. Additionally, the study employed a combination of strategic manual searches and electronic database queries. While this approach aimed to enhance the study's scope, it introduced subjective judgments during the initial selection and detailed examination of

texts. Although a collaborative approach was used to reconcile scholarly disagreements, the risk of interpretive bias in selecting and analyzing studies persists. In instances of incomplete or combined data sets, direct contact was made with original authors to obtain specific data relevant to MCI or AD cohorts. While this proactive step was intended to improve data accuracy and completeness, it may be subject to response bias depending on the original authors' willingness and ability to provide data.

The analytical framework of the study included both quantitative and qualitative synthesis of outcomes, applying quantitative synthesis in cases of clinical and methodological consistency and qualitative summaries in cases of notable heterogeneity. While this comprehensive approach facilitated a nuanced understanding of the measurement properties, it also introduced limitations in the uniform interpretation and integration of heterogeneous data sets, potentially impacting the overall methodological soundness of the conclusions drawn. In summary, this study provides valuable insights into the methodological quality of cognitive assessment tools for MCI and AD, but the outlined limitations necessitate cautious extrapolation of the results and underscore the need for further methodologically sound research in this area.

Additionally, none of the assessment instruments consistently excelled across all measurement properties, indicating variable quality. To recapitulate, it is crucial to emphasize the necessity for continued and methodologically robust research in the field of cognitive assessment tools for MCI and AD diagnosis. Our study highlights a spectrum of methodological qualities across various cognitive assessment instruments, ranging from moderate to excellent. This variability underscores the critical need for future research to focus on refining these tools to enhance their sensitivity, specificity, reliability, and applicability in diverse clinical and research settings. Potential research avenues should aim to address the limitations identified in this study, such as exploring cognitive assessment tools in languages beyond English and Spanish to incorporate a broader range of cultural and linguistic contexts. Additionally, further investigation is needed to understand the efficacy of these tools in different stages of cognitive impairment and across various demographic groups, including age, education level, and cultural backgrounds. There is also a pressing need for longitudinal studies to assess the long-term reliability and validity of these cognitive assessment tools in predicting the progression of MCI to AD. Such studies would provide invaluable insights into the predictive power of these tools, thereby enhancing their clinical utility. Furthermore, integrating technological advancements like digital assessments and machine learning could improve the accuracy and efficiency of cognitive evaluations. Developing sophisticated tools that integrate seamlessly into clinical practices to deliver precise diagnostics is a critical research direction. Ultimately, this review sets the stage for future efforts aimed at creating refined, reliable, and universally applicable cognitive assessment tools that facilitate the early and accurate diagnosis of MCI and AD, leading to better patient outcomes.

Conclusion

This study sheds light on the varied effectiveness of cognitive assessment tools for diagnosing MCI and AD, underscoring the critical need for selecting tools that align with specific diagnostic and monitoring requirements. We find that the MMSE-2, although widely used, shows moderate sensitivity, particularly in detecting early-stage cognitive impairments. In contrast, the culturally adaptive T-CMMSE proves more effective in linguistically and culturally diverse groups. The HK-MoCA-A1 and A2 are specifically effective for the Hong Kong population, highlighting the significance of cultural relevance. The MoCA stands out for its ability to detect early cognitive decline across various demographics, demonstrating its versatility.

However, the study is limited by its demographic scope and the range of tools assessed. Future research should aim for greater inclusivity and broader evaluation of tools to enhance the generalizability of the findings. Our results support a tailored approach to selecting cognitive assessment tools, tailored to the specific profiles and needs of individual patients, thus improving diagnostic accuracy and the effectiveness of treatments in managing MCI and AD. This research represents a significant advance in furthering the understanding of cognitive assessment in neurodegenerative disorders, guiding us towards improved patient care and outcomes in addressing these intricate conditions.

ORCID iDs

Xia Peng https://orcid.org/0000-0003-2514-2980
Shi-Fen Xu https://orcid.org/0000-0001-5595-3480

Statements and declarations

References

1. Manly JJ, Jones RN, Langa KM, et al. Estimating the prevalence of dementia and mild cognitive impairment in the US: the 2016 health and retirement study harmonized cognitive assessment protocol project. *JAMA Neurol* 2022; 79: 1242–1249.

2. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19: 539–549.

3. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN Guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27: 1147–1157.

4. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27: 1171–1179.

5. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN Methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018; 27: 1159–1170.

6. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; 6: e1000100. 2009/07/22.

7. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 2021; 88: 105906.

8. Baek MJ, Kim K, Park YH, et al. The validity and reliability of the Mini-mental state Examination-2 for detecting mild cognitive impairment and Alzheimer's disease in a Korean population. *PLoS One* 2016; 11: e0163792.

9. Wong SS and Fong KN. Reliability and validity of the telephone version of the Cantonese Mini-mental state examination (T-CMMSE) when used with elderly patients with and without dementia in Hong Kong. *Int Psychogeriatr* 2009; 21: 345–353.

10. Keskinoglu P, Ucku R, Yener G, et al. Reliability and validity of revised Turkish version of Mini mental state examination (rMMSE-T) in community-dwelling educated and uneducated elderly. *Int J Geriatr Psychiatry* 2009; 24: 1242–1250.

11. El-Hayeck R, Baddoura R, Wehbe A, et al. An Arabic version of the Mini-mental state examination for the Lebanese population: reliability, validity, and normative data. *J Alzheimers Dis* 2019; 71: 525–540.

12. Aguilar-Navarro SG, Mimenza-Alvarado AJ, Palacios-Garcia AA, et al. Validity and reliability of the Spanish version of the Montreal Cognitive Assessment (MoCA) for the detection of cognitive impairment in Mexico. *Rev Colomb Psiquiatr (Engl Ed)* 2018; 47: 237–243.

13. Mougias AA, Christidi F, Kiosterakis G, et al. Dealing with severe dementia in clinical practice: a validity and reliability study of severe Mini-mental state examination in Greek population. *Int J Geriatr Psychiatry* 2018; 33: 1236–1242.

14. Wong A, Yiu S, Nasreddine Z, et al. Validity and reliability of two alternate versions of the Montreal cognitive assessment (Hong Kong version) for screening of mild neurocognitive disorder. *PLoS One* 2018; 13: e0196344.

15. Gierus J, Mosiolek A, Koweszko T, et al. The Montreal cognitive assessment as a preliminary assessment tool in general psychiatry: validity of MoCA in psychiatric patients. *Gen Hosp Psychiatry* 2015; 37: 476–480.

16. Bezdicek O, Cervenkova M, Moore TM, et al. Determining a short form Montreal Cognitive Assessment (s-MoCA) Czech version: validity in mild cognitive impairment Parkinson's disease and cross-cultural comparison. *Assessment* 2020; 27: 1960–1970.

17. Gupta M, Gupta V, Nagar Buckshee R, et al. Validity and reliability of hindi translated version of Montreal cognitive assessment in older adults. *Asian J Psychiatr* 2019; 45: 125–128.

18. Karim MA and Venkatachalam J. Construct validity and psychometric properties of the Tamil (India) version of Montreal Cognitive Assessment (T-MoCA) in elderly. *Int J Gerontol* 2022; 16: 365–369.

19. Rahman TT and El Gaafary MM. Montreal Cognitive assessment arabic version: reliability and validity prevalence of mild cognitive impairment among elderly attending geriatric clubs in Cairo. *Geriatr Gerontol Int* 2009; 9: 54–61.

20. Janelidze M, Mikeladze N, Bochorishvili N, et al. Validity of the Georgian Montreal cognitive assessment for the screening of mild cognitive impairment and dementia. *Am J Alzheimers Dis Other Demen* 2017; 32: 36–40.

21. Sarasola D, de Lujan-Calcagno M, Sabe L, et al. Validity of the Spanish version of the Addenbrooke's cognitive examination for the diagnosis of dementia and to differentiate Alzheimer's disease and frontotemporal dementia. *Rev Neurol* 2005; 41: 717–721.

22. Rashedi V, Foroughan M, Nazari H, et al. Validity and reliability of the Persian version of general practitioner assessment of cognition (P-GPCOG). *Aging Ment Health* 2019; 23: 961–965.

23. Morita A, O'Caoimh R, Murayama H, et al. Validity of the Japanese version of the quick mild cognitive impairment screen. *Int J Environ Res Public Health* 2019; 16: 917.

24. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21: 651–657.

25. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34–42.

26. Freitas S, Simoes MR, Maroco J, et al. Construct validity of the Montreal Cognitive Assessment (MoCA). *J Int Neuropsychol Soc* 2012; 18: 242–250.

27. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017; 358: j4008.

28. Vos SJ, Verhey F, Frolich L, et al. Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *Brain* 2015; 138: 1327–1338.

29. Jia J, Zhou A, Wei C, et al. The prevalence of mild cognitive impairment and its etiological subtypes in elderly Chinese. *Alzheimers Dement* 2014; 10: 439–447.

30. Shigemori K, Ohgi S, Okuyama E, et al. The factorial structure of the Mini-Mental State Examination (MMSE) in Japanese dementia patients. *BMC Geriatr* 2010; 10: 36.

31. Potts C, Richardson J, Bond RB, et al. Reliability of Addenbrooke's Cognitive Examination III in differentiating between dementia, mild cognitive impairment and older adults who have not reported cognitive problems. *Eur J Ageing* 2022; 19: 495–507.