Diagnostic accuracy, reliability, and construct validity of the German quick mild cognitive impairment screen

Patrick Manser[1]* and Eling D. de Bruin[1,2,3]

Abstract

Background Early detection of cognitive impairment is among the top research priorities aimed at reducing the global burden of dementia. Currently used screening tools have high sensitivity but lack specificity at their original cut-off, while decreasing the cut-off was repeatedly shown to improve specificity, but at the cost of lower sensitivity. In 2012, a new screening tool was introduced that aims to overcome these limitations – the Quick mild cognitive impairment screen (Qmci). The original English Qmci has been rigorously validated and demonstrated high diagnostic accuracy with both good sensitivity and specificity. We aimed to determine the optimal cut-off value for the German Qmci, and evaluate its diagnostic accuracy, reliability (internal consistency) and construct validity.

Methods We retrospectively analyzed data from healthy older adults (HOA; $n = 43$) and individuals who have a clinical diagnosis of 'mild neurocognitive disorder' (mNCD; $n = 37$) with a biomarker supported characterization of the etiology of mNCD of three studies of the 'Brain-IT' project. Using Youden's Index, we calculated the optimal cut-off score to distinguish between HOA and mNCD. Receiver operating characteristic (ROC) curve analysis was performed to evaluate diagnostic accuracy based on the area under the curve (AUC). Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. Reliability (internal consistency) was analyzed by calculating Cronbach's α. Construct validity was assessed by analyzing convergent validity between Qmci-G subdomain scores and reference assessments measuring the same neurocognitive domain.

Results The optimal cut-off score for the Qmci-G was ≤ 67 (AUC = 0.96). This provided a sensitivity of 91.9% and a specificity of 90.7%. The PPV and NPV were 89.5% and 92.9%, respectively. Cronbach's α of the Qmci-G was 0.71 (CI$_{95\%}$ [0.65 to 0.78]). The Qmci-G demonstrated good construct validity for subtests measuring learning and memory. Subtests that measure executive functioning and/or visuo-spatial skills showed mixed findings and/or did not correlate as strongly as expected with reference assessments.

Conclusion Our findings corroborate the existing evidence of the Qmci's good diagnostic accuracy, reliability, and construct validity. Additionally, the Qmci shows potential in resolving the limitations of commonly used screening tools, such as the Montreal Cognitive Assessment. To verify these findings for the Qmci-G, testing in clinical environments and/or primary health care and direct comparisons with standard screening tools utilized in these settings are warranted.

*Correspondence:
Patrick Manser
patrick.manser@hest.ethz.ch

Full list of author information is available at the end of the article

Introduction

Background

Three of the top ten current research priorities aimed at reducing the global burden of dementia are centered around the prevention, early identification, and mitigation of dementia risk factors [1]. In this context, it is imperative that significant emphasis is placed on improving the timely and accurate detection and diagnosis of mild and major neurocognitive disorder (mNCD and MNCD; formerly referred to as 'mild cognitive impairment' and 'dementia' [2–6]) to facilitate early interventions as part of the secondary prevention of mNCD [1].

Neurocognitive disorders are currently largely underdiagnosed, with a global pooled prevalence of undetected MNCDs of 61.7% in middle- and high-income countries. There are various possible explanations for this phenomenon. For instance, primary care physicians and health professionals may still consider cognitive difficulties a common trait of normal aging rather than a disability that necessitates specialized attention and support. As a result, they may hesitate to refer these patients to memory clinics for a clinical diagnosis [7]. The widespread and consistent use of a validated and recognized screening tool would undoubtedly improve the ability to detect individuals with suspected NCDs who should be referred for a clinical diagnosis [7], which is consistent with the majority of currently available clinical practice guidelines for the diagnosis and treatment of mNCD [8]. However, recommendations by the United States Preventive Services Task Force in 2020 and the Canadian Task Force on Preventive Health in 2016 do not suggest screening for cognitive impairment or dementia in asymptomatic older adults due to the lack of evidence demonstrating its advantages as well as potentially high rate of false-positive screens [9, 10].

The most frequent screening tools for individuals with suspected NCDs utilized in clinical practice and research [8, 11–13] include the Mini-Mental State Examination (MMSE) [14] and the Montreal Cognitive Assessment (MoCA) [15]. The MoCA was found the most common and preferable tool for screening of mNCD [13] and is superior to the MMSE in the detection of mNCD [16]. However, while the initially proposed cut-off (<26 points) [15] has shown good sensitivity for discriminating mNCD from healthy older adults (HOA) [15, 17] with a pooled sensitivity of 93.7% [18], this cut-off was repeatedly found to have low specificity [17] (pooled specificity $= 58.8\%$) [18]. Similar findings have been obtained for the original cut-off for the German MoCA, with a sensitivity and specificity of 86% and 63%, respectively [19].

Decreasing the cut-off was repeatedly shown to improve specificity, but at the cost of lower sensitivity. Therefore, it was recommended to adjust the utilized cut-off scores based on the preferred prioritization of sensitivity or specificity [18], which has been thoroughly investigated in the German MoCA [19]. Thomann et al. (2020) concluded that *"using two separate cut-offs for the MoCA combined with scores in an indecisive area enhances the accuracy of cognitive screening"* [19]. Alternatively, more robust and accurate screening tools should be developed [18].

In 2012, a new screening tool was introduced that aims to overcome these limitations – the Quick mild cognitive impairment screen (Qmci) [20–22]. In comparison to the MMSE and MoCA, the Qmci has a more detailed scoring system and includes a logical memory task, which allows it to detect subtle cognitive changes and avoid ceiling effect [23]. The original English Qmci was shown to accurately discriminate between individuals with normal cognitive functioning ($n=623$), mNCD ($n=147$), and MNCD ($n=165$) [24]. In addition, the Qmci has undergone successful validation in multiple languages, including Chinese [25], Dutch [26], Greek [27, 28], Japanese [29], Persian [30], Taiwanese [31], and Turkish [32]. However, the optimal cut-off for discriminating between mNCD and HOA as well as the diagnostic accuracy of the German version of the Qmci (Qmci-G) have not yet been scientifically determined and validated. These investigations are required for the Qmci-G to be used in German-speaking countries.

Objectives

The primary objective of this study was to determine the optimal cut-off value for the Qmci-G and to evaluate its diagnostic accuracy. As secondary objectives, we assessed the reliability (internal consistency) of the Qmci-G and explored the construct validity of the Qmci-G in older adults who have mNCD.

Methods

Study Design and participants

This study retrospectively analyzed data of three studies of the 'Brain-IT' project, namely baseline assessments of a cross-sectional study which included assessments of the Qmci-G in HOA [33] and two intervention studies that assessed the feasibility [34] and effectiveness ([35, 36]) of a novel technology-supported training concept for the secondary prevention of mNCD. The study was reported according to the Standards for Reporting of Diagnostic

Accuracy Studies guidelines and elaboration paper [37, 38] (checklist see supplementary file 1).

In the cross-sectional study, HOA (healthy based on self-report and ≥60 years) were recruited between January 2021 and June 2021 in collaboration with healthcare institutions in the larger area of Zurich by handing out leaflets to interested persons. In the two intervention studies, older adults who have mNCD were recruited between July 2021 and October 2023 in collaboration with (memory) clinics in the larger area of Zurich and St. Gallen. All suitable patients were identified through medical records and patient registries at these (memory) clinics, or through recent clinical diagnostics. For this study, we only analyzed data of all participants who have a biomarker supported characterization of the etiology of mNCD in addition to the clinical diagnosis of 'mild neurocognitive disorder' according to International Classification of Diseases 11th Revision (ICD-XI) [40] or the latest Diagnostic and Statistical Manual of Mental Disorders 5th Edition (DSM-5®) [7]. Besides these inclusion criteria for the characterization of the population (HOA and mNCD), the same eligibility criteria were used in all three studies (for the full list of eligibility criteria, refer to Table 1).

The first author (PM) was responsible for the design, implementation, conduct, and analysis of all three of these studies under the supervision of EdB. He trained each involved study investigator for all study procedures according to Guidelines of Good Clinical Practice and in line with detailed working instructions and was in charge of methodological standards and quality of data collection under the supervision of EdB. The same working instructions were followed in all three studies. These detailed working instructions standardized all measurement procedures and instructions of participants to minimize bias during assessment of all outcome measures.

Outcomes

Primary outcome: Qmci

As primary outcome, data of the Qmci-G [20, 21] was used. The Qmci consists of six subtests: orientation (10 points), registration (5 points), clock drawing (15 points), delayed recall (20 points), verbal fluency (20 points), and logical memory (30 points) and is scored out of a maximum of 100 points [21, 22]. It was administered and evaluated according to published guidelines [21]. According to these guidelines, administration and scoring of the Qmci should not exceed 5 min [21].

Secondary outcomes

As secondary outcomes, data of assessments of the neurocognitive domains of learning and memory, executive functions, and visuo-spatial skills was used. For learning and memory, data of the German version of the subtest 'logical memory' of the Wechsler Memory Scale – fourth edition (WMS-IV-LM) [39, 40] was used. For executive functions, we considered data for working memory (i.e. using a computerized version of the Digit Span Backward test (Psychology Experiment Building Language (PEBL) - Digit Span Backward (PEBL-DSB)) [41–43], cognitive flexibility (i.e. using a computerized version of the Trail Making Test – Part B (PEBL-TMT-B) [41, 43]), and planning abilities (i.e. using the HOTAP picture-sorting test part A (HOTAP-A) [44]). For visuo-spatial skills, we considered data from a computerized Mental Rotation Task (PEBL-MRT) [41, 43, 45] that is based on the classic mental rotation task by Shepard and Metzler [46]. All assessments were administered and evaluated in accordance with published guidelines or detailed working instructions. For further information on these assessments, please refer to the study protocol of our RCT [36].

Other outcomes

Baseline factors were collected through demographic data including age, sex, height, weight, body mass index (BMI), years of education, and (for participants who have

Table 1 Description of all eligibility criteria

Inclusion criteria:	Exclusion criteria:
Participants fulfilling all the following inclusion criteria were eligible:	The presence of any of the following criteria led to exclusion:
• (1 = mNCD) clinical diagnosis of 'mild neurocognitive disorder' according to International Classification of Diseases 11th Revision (ICD-XI) [6] or the latest Diagnostic and Statistical Manual of Mental Disorders 5th Edition (DSM-5®) [5]) AND biomarker supported characterization of the etiology of mNCD OR (2 = HOA) healthy (based on self-report) older adults (≥60 years) • German speaking • able to stand for at least 10 min without assistance	• mobility impairments (i.e., gait, balance) that prevent experiment participation • presence of additional, clinically relevant (i.e., acute and/or symptomatic) neurological disorders (i.e., epilepsy, stroke, multiple sclerosis, Parkinson's disease, brain tumors, or traumatic disorders of the nervous system) • presence of any other unstable or uncontrolled diseases (e.g., uncontrolled high blood pressure, progressing or terminal cancer)

Abbreviations: HOA, healthy older adults; mNCD, mild neurocognitive disorder

mNCD) classification of etiology of mNCD (biomarker supported).

Statistics

Statistical analysis was done after data collection was completed using R (4.3.1 GUI 1.79 Big Sur Intel build) in line with RStudio (Version 2023.06.2+561). Data was reported as means±standard deviations for data fulfilling all the assumptions that would subsequently justify parametric statistical analyses. In case these assumptions were not met, medians (interquartile ranges) were reported. First, descriptive statistics were computed for all outcome variables [47–49]. The normality of the data was checked using the Shapiro-Wilk test. For all demographic variables, between-group differences (i.e., HOA and older adults who have mNCD) were tested using an independent t-test or Mann–Whitney U-test in case the data were not normally distributed. Between-group differences in categorical variables were tested using Fisher's exact test. To discover whether the between-group differences were substantive, Pearson's r effect sizes were calculated [49, 50] and interpreted to be small ($0.1 \leq r < 0.3$), medium ($0.3 \leq r < 0.5$) or large ($r > 0.5$) [51]. The level of significance was set at $p \leq 0.05$ (one-sided).

Optimal cut-off value and diagnostic accuracy of the Qmci-G

The optimal cut-off score for discriminating between HOA and older adults who have mNCD was calculated using Youden's Index [52] in the OptimalCutpoints package [53]. Receiver operating characteristic (ROC) curve analysis was done using the pROC package [54] and used to assess diagnostic accuracy based on the area under the curve (AUC). The resulting AUC value was interpreted to represent poor ($0.60 \leq |AUC| < 0.70$), fair ($0.70 \leq |AUC| < 0.80$), good ($0.80 \leq |AUC| < 0.9$), and excellent ($|AUC| \geq 0.90$) discriminatory ability [55, 56]. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for the optimal cut-off score.

Reliability (internal consistency) of the Qmci-G

Cronbach's α was calculated to investigate the internal consistency of the Qmci-G [49]. The degree of consistency was interpreted according to the categorization for Cronbach's α defined in [57]. Cronbach's α≥0.70 was set as the criterion for "adequate" internal consistency [57].

Construct validity of the Qmci-G in older adults who have mNCD

The Qmci-G subtest 'orientation' is mainly intended to distinguish between individuals who have mNCD and MNCD and therefore has limited discriminatory power between HOA and individuals who have mNCD due to ceiling effects [58, 59]. Therefore, assessment of construct validity focused on the remaining subtests of the Qmci in this study.

Construct validity of the Qmci-G was assessed by analyzing convergent validity between the Qmci-G subdomain scores and reference assessments measuring the same neurocognitive domain according to Sachdev et al. 2014 [3]. It was hypothesized that there is a significant strong positive correlation between: (alternative hypothesis number 1 ($H_{A,1}$):) Qmci-G subtest 'registration' and reference assessments for auditory learning and memory; ($H_{A,2}$:) Qmci-G subtest 'clock drawing' and reference assessments for executive functions/visuo-spatial skills; ($H_{A,3}$:) Qmci-G subtest 'recall' and reference assessments for auditory learning and memory; ($H_{A,4}$:) Qmci-G subtest 'logical memory' and reference assessments for auditory learning and memory; ($H_{A,5}$:) Qmci-G subtest 'verbal fluency' and reference assessments for executive functions.

One-sided bivariate correlation analyses were performed for the neurocognitive domains of (1) learning and memory (i.e., between Qmci-G subscores 'registration', 'delayed recall' as well as 'logical memory' and WMS-IV-LM 1 and 2 [39, 40] for auditory learning and memory; (2) executive functions (i.e., Qmci-G subscores 'verbal fluency' and PEBL-DSB [41–43], PEBL-TMT-B [41, 43]), as well as HOTAP-A [44]; and (3) and combined executive functions/visuo-spatial skills (i.e., Qmci-G subscore 'clock drawing' and PEBL-DSB [41–43], PEBL-TMT-B [41, 43], as well as PEBL-MRT [41, 43, 45, 46]). Pearson's correlation coefficients (r) were computed for datasets adhering to assumptions for parametric analyses and Spearman's rank correlation coefficients (r_s) for datasets violating assumptions for parametric analyses. 95% confidence intervals ($CI_{95\%}$) were calculated using the R-package 'ci_cor'. For spearman correlation coefficients, we used bootstrap $CI_{95\%}$ with the bias-corrected and accelerated method, 999 bootstrap resamples and 1,000 seeds. The resultant correlation coefficients were interpreted as weak ($0.1 \leq |r_{(s)}| < 0.3$), moderate ($0.3 \leq |r_{(s)}| < 0.5$) or strong ($|r_{(s)}| \geq 0.5$) correlation [49, 51]. The alternative hypotheses (i.e., convergent validity between the Qmci-G subdomain scores and reference assessments measuring the same neurocognitive domain) were considered confirmed in case of: (a) a significant ($p \leq 0.05$, one-tailed) positive correlation between the Qmci-G subdomain score and the corresponding reference assessment, and (b) a correlation coefficient of $|r_{(s)}| \geq 0.4$ [60].

Sample size justification

In this study, we did not perform an a-priori sample size calculation as we analyzed existing datasets from studies conducted as part of the 'Brain-IT' project. This approach

is supported by various factors and aligns with our research objectives.

Extensive data on the diagnostic accuracy of the original English Qmci is available. As summarized in the introduction, the original Qmci was shown to accurately discriminate between individuals with normal cognitive functioning, mNCD, and MNCD [24] with high diagnostic accuracy, sensitivity, and specificity. In addition, the Qmci has undergone successful validation in multiple languages, including Chinese [25], Dutch [26], Greek [27, 28], Japanese [29], Persian [30], Taiwanese [31], and Turkish [32]. These studies were adequately powered according to a priori sample size calculations. The robustness of our retrospective data analysis is validated due to the dataset's similar sample size to most of these studies.

In this study, our aim was to corroborate and build upon these research findings referenced above. To this end, we primarily aimed to ensure that our study population was representative in terms of demographic characteristics as well as descriptive statistics of the Qmci to optimize generalizability of our findings and critically discuss whether this was successful in section 'Discussion – Generalizability of the Findings'.

Results

Descriptive statistics of Study Population

The descriptive statistics of the study population are summarized in Table 2. There were no adverse events related to any of the study's measurements.

Table 2 Demographic characteristics of the Study Population;

	Group 1: HOA (n = 43)	Group 2: mNCD (n = 37)	Between-Group Difference		
			test statistics[1]	*p*-value[2]	effect size [CI$_{95\%}$] [3]
Age [years]	67.0 (7.0)	77.0 (10.0)	W = 423	< 0.001*	r_s = -0.402 [-0.571 to -0.120]
Sex [% females]	58.1	32.4	N/A	0.026*	OR = 2.9 [1.2 to 7.2]
Body mass index [kg·m^{-2}]	23.7 ± 3.0	23.2 (3.9)	W = 829	0.750	r_s = -0.036 [-0.253 to 0.186]
Years of education [years]	15.9 ± 4.3	14.9 ± 4.0	t = 0.8	0.416	*r* = 0.107 [-0.142 to 0.342]
Etiology of mNCD:					
mNCD due to Alzheimer's Disease	N/A	n = 25 (67.6%)	N/A	N/A	N/A
mild frontotemporal NCD	N/A	n = 3 (8.1%)	N/A	N/A	N/A
mNCD with Lewy Bodies	N/A	n = 1 (2.7%)	N/A	N/A	N/A
mild vascular NCD	N/A	n = 6 (16.2%)	N/A	N/A	N/A
unclear / not (yet) determined	N/A	n = 2 (5.4%)	N/A	N/A	N/A
Qmci-G total score	78.8 ± 8.6	54.5 ± 13.0	t = 10.3	< 0.001*	*r* = -0.799 [-0.872 to -0.691]
Qmci-G subscore 'orientation'	10.0 (0.0)	10.0 (2.0)	W = 1,120	< 0.001*	r_s = -0.489 [-0.640 to -0.302]
Qmci-G subscore 'registration'	5.0 (0.5)	4.0 (2.0)	W = 1,233	< 0.001*	r_s = -0.517 [-0.662 to -0.336]
Qmci-G subscore 'clock drawing'	15.0 (0.0)	13.0 (2.0)	W = 1,400	< 0.001*	r_s = -0.690 [-0.790 to -0.555]
Qmci-G subscore 'recall'	20.0 (4.0)	12.0 (8.0)	W = 1,297	< 0.001*	r_s = -0.562 [-0.696 to -0.391]
Qmci-G subscore 'fluency'	12.0 ± 2.4	7.8 ± 2.3	t = 8.0	< 0.001*	*r* = -0.673 [-0.776 to -0.534]
Qmci-G subscore 'logical memory'	20.4 ± 5.0	9.7 ± 6.2	t = 8.4	< 0.001*	*r* = -0.710 [-0.807 to -0.577]

Data is reported as means ± standard deviations for data fulfilling all the assumptions to justify parametric statistical analyses. In case these assumptions were not met, medians (interquartile ranges) are reported

[1] t-statistics for the between-group differences tested with an independent t-test or Mann-Whitney U test in case the data are not normally distributed;

[2] *p*-values for the between-group differences tested with an independent t-test or Mann-Whitney U test in case the data are not normally distributed, or Fisher's exact test for categorical variables

[3] effect size estimates for the between-group differences tested with an independent t-test (effect size Pearson r) or Mann-Whitney U test (effect size Spearman rho (r_s)) in case the data are not normally distributed, or Fisher's exact test for categorical variables (odds ratio)

* = significant at *p* < 0.05.

Abbreviations: HOA, healthy older adults; mNCD, mild neurocognitive disorder; Qmci-G, German version of the Quick mild cognitive impairment screen
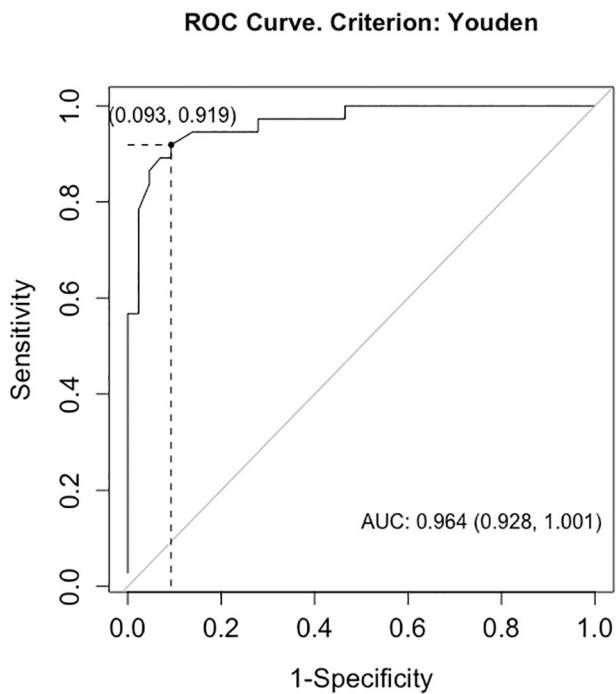
ROC Curve. Criterion: Youden

Fig. 1 Receiver operating characteristic (ROC) curve with the optimal cut-off score for discriminating between healthy older adults and older adults who have mNCD calculated using Youden's Index. Abbreviations: AUC, area under curve; ROC, receiver operating characteristic

Optimal cut-off value and diagnostic accuracy of the Qmci-G

Using Youden's Index, the optimal cut-off score for the Qmci-G to discriminate between HOA and older adults who have mNCD was ≤67 (AUC=0.96; 95% confidence interval (CI$_{95\%}$): 0.93, 1.00). This provided a sensitivity of 91.9% and a specificity of 90.7% (see Fig. 1). The PPV and NPV were 89.5% and 92.9%, respectively.

Reliability (internal consistency) of the Qmci-G

Cronbach's α of the Qmci-G was 0.71 (CI$_{95\%}$ [0.65 to 0.78]).

Construct validity of the Qmci-G in older adults who have mNCD

The $r_{(s)}$ and *p*-values for the correlation between the Qmci-G subtest scores and the scores of the corresponding reference assessments for each hypothesis are summarized in Table 3. The Qmci-G subtests assessing learning and memory showed a significant and strong correlation with the reference assessments. Subtests that measure executive functioning and/or visuo-spatial skills showed mixed findings and/or did not correlate as strong as expected with reference assessments.

Table 3 Bivariate correlation analyses between the German version of the quick mild cognitive impairment screen (Qmci-G) subtest scores and the scores of the corresponding reference assessments

Construct	Statistics	
	p-value[1]	$r_{(s)}$ [CI$_{95\%}$][1]
Learning and Memory:		
Qmci-G subscore 'registration' - WMS-IV-LM 1 (score)	0.013*	0.366 [-0.007 to 0.612]
Qmci-G subscore 'registration'- WMS-IV-LM 2 (score)	< 0.001*	0.508 [0.207 to 0.724
Qmci-G subscore 'delayed recall' - WMS-IV-LM 1 (score)	< 0.001*	0.516 [0.183 to 0.742]
Qmci-G subscore 'delayed recall' - WMS-IV-LM 2 (score)	< 0.001*	0.604 [0.284 to 0.802]
Qmci-G subscore 'logical memory' - WMS-IV-LM 1 (score)	< 0.001*	0.721 [0.517 to 0.847]
Qmci-G subscore 'logical memory' - WMS-IV-LM 2 (score)	0.005*	0.414 [0.104 to 0.651]
Executive Functions:		
Qmci-G subscores 'verbal fluency'- PEBL-DSB (score)	0.038*	0.312 [-0.035 to 0.592]
Qmci-G subscores 'verbal fluency'- PEBL-TMT-B (time)	0.125	0.200 [-0.143 to 0.500]
Qmci-G subscores 'verbal fluency'- HOTAP-A (combi-score)	0.149	0.178 [-0.159 to 0.479]
Executive Functions/Visuo-spatial Skills:		
Qmci-G subscores 'clock drawing' - PEBL-DSB (score)	0.001*	0.504 [0.136 to 0.748]
Qmci-G subscores 'clock drawing' - PEBL-TMT-B (time)	0.004*	0.437 [0.098 to 0.688]
Qmci-G subscores 'clock drawing' - PEBL-MRT (points)	0.170*	0.178 [-0.171 to 0.492]
Qmci-G subscores 'clock drawing' - PEBL-MRT (reaction time)	0.051	0.300 -0.067 to 0.582]

[1] Pearson's correlation coefficients (r) for datasets adhering to assumptions for parametric analyses, Spearman's rank correlation coefficients (r_s) for datasets violating assumptions for parametric analyses, and *p*-values

* = significant at *p* < 0.05

Abbreviations: CI$_{95\%}$, 95% confidence interval; DSB, Digit Span Backward; HOTAP-A, HOTAP picture-sorting test part A; MRT, Mental Rotation Task; PEBL, Psychology Experiment Building Language; Qmci, Quick mild cognitive impairment screen; TMT-B, Trail Making Test – Part B; WMS-IV-LM, subtest 'logical memory' of the Wechsler Memory Scale – fourth edition

Discussion

This study determined the optimal cut-off value for the Qmci-G, and evaluated its diagnostic accuracy, reliability (internal consistency) and construct validity. The key findings of this study are that the Qmci-G demonstrated (1) excellent discriminatory ability between HOA and older adults who have mNCD at its optimal cut-off score of ≤67 points; (2) adequate internal consistency; and (3) good construct validity for subtests measuring learning and memory. However, subtests that measure executive functioning and/or visuo-spatial skills showed mixed findings and/or did not correlate as strongly as expected with reference assessments.

Diagnostic accuracy of the Qmci

The excellent discriminatory ability of the Qmci-G between HOA and mNCD found in this study is consistent with extensive data on good diagnostic accuracy of the Qmci [11, 24] and, therefore, corroborates the available evidence of the original English Qmci (AUC=0.84, optimal cut-off (Youden Index)≤67, Sensitivity=77%, Specificity=75%) [24]. In addition, our findings are consistent with pooled data of 2019 for the Qmci demonstrating an AUC of 0.84 [11], a sensitivity between 77% [11] to 82% [23], and a specificity of 79% [11] to 82% [23] at a given cut-off score (i.e., the recommended cut-off score or, in case several sensitivity/specificity pairs were presented in the original studies analyses in these systematic reviews and meta-analyses, the cut-off score that was described as optimal by the respective authors or produced the largest AUC). Finally, our findings are also consistent with more recent validation studies of the Qmci in other translations, including Greek (AUC=0.79 [28] and AUC=0.76 [27]), Japanese (Sensitivity=94%, Specificity=72%) [29], Persian (AUC=0.80) [30], Taiwanese (AUC=0.89) [31], and Turkish (AUC=0.80) [32]. The considerably higher AUC as compared to previous publications may be attributed to several characteristics of our analysis pertaining to the recruitment and characteristics of the participant sample under investigation. These are discussed in greater detail in the sections "Generalizability of the Findings" and "Strengths and Limitations".

More importantly, there is evidence from a systematic review [23] and a meta-analysis [11] demonstrating that the Qmci has comparable [11] to superior [23] accuracy, sensitivity, and specificity compared to the standardized MMSE and the MoCA in detecting cognitive impairment [11, 23]. This finding holds significance for research and clinical practice, because the MMSE and MoCA are the most widely used screening instruments for mNCD [8, 12] and are known to have high sensitivity but low specificity at its original cut-off [16–18], also in the German MoCA (specificity=63% at original cut-off) [19], while decreasing the cut-off was repeatedly shown to improve specificity, but at the cost of lower sensitivity [18]. In contrast, our results indicate both high sensitivity and specificity. This may be explained because the Qmci was developed on basis of the AB Cognitive Screen 135 [58] by reweighting its scoring and introducing a logical memory task with the aim to increase sensitivity and particularly specificity to detect mNCD [61]. This appears successful, since the logical memory subtest exhibited the highest accuracy of all Qmci subtests in discriminating between HOA and older adults with mNCD, as evidenced by an AUC of 0.80 [59]. Moreover, pooled sensitivity and specificity estimates of the Qmci were shown to not significantly differ from comprehensive cognitive assessments [11], such as the Addenbrooke's Cognitive Examination Revised (ACE-R) [62], the Consortium to Establish a Registry for Alzheimer's Disease Battery (CERAD) total score [63]. However, the Qmci has a substantially shorter administration time. According to published guidelines, administration and scoring should not exceed 5 min [21], which aligns with the median administration times of 4.5 min [64] to 5 min [11] reported in the literature. In contrast, the MoCA has a median administration time of 9.5 min [64] to 12 min [11], the ACE-R takes 12 to 20 min [11, 62] and the CERAD takes 20 to 30 min [11, 65]. This substantially shorter administration time, coupled with comparable [11] or even marginally superior [23] diagnostic accuracy suggests that the Qmci has potential as a means of assessing patients who present with cognitive complaints in primary care [11] and thereby allow more widespread and consistent use of a validated screening tool. This could ultimately aid in the early detection of individuals with suspected mNCD, facilitating their referral for clinical diagnosis [7], and supporting the implementation of interventions as part of the secondary prevention of mNCD, all of which are currently among the top ten research priorities in reducing the global burden of dementia [1].

Reliability (internal consistency) and construct validity of the Qmci

The internal consistency of the Qmci-G was found adequate (Cronbach's α≥0.70 [57]). This is consistent with previous research as evidenced by Cronbach's α values of 0.71 [28], 0.81 [30], 0.85 [31], 0.95 [20] and indicates that the subtests of the Qmci consistently and reliably assess the same underlying construct or concept (global cognitive functioning). With regards to the construct validity of the Qmci, previous studies have only analyzed the convergent validity between the Qmci total score and reference assessments [21, 28, 29, 31, 32, 66, 67]. These studies have shown significant strong positive correlations with the MoCA [31, 32] and significant weak [66], moderate [29], or strong [28, 31] positive correlations with the MMSE. In addition, the Qmci also showed significant

strong correlations with a detailed neuropsychological battery (the standardized Alzheimer's Disease Assessment Scale - cognitive subscale (SADAS-cog) [68, 69]) as well as the Clinical Dementia Rating scale [70] and was shown to be responsive to change over time [67]. This supports the construct validity of the Qmci and indicates that it could be used as a substitute for more comprehensive neuropsychological assessments in clinical trials [67].

As we were limited by the data available in this retrospective data analysis, we were unable to confirm these findings with the Qmci-G, due to unavailability of data from reference assessments for global cognitive functioning. However, we demonstrated construct validity for the Qmci subtest measuring learning and memory, as evidenced by significant and mostly strong correlations with clinically validated reference assessments for learning and memory. This is an important finding, as 55% of the maximum total score is allocated to the Qmci subtests measuring learning and memory (i.e., 'registration' (maximum 5 points), 'recall' (maximum 20 points), and 'logical memory' (maximum 30 points) [20, 21]. It is well known that the first line of prediction of mNCD is assessment of the neurocognitive domain of learning and memory [71]. Therefore, one potential explanation for why the Qmci may be more specific than the MoCA is due to the MoCA's lesser emphasis on learning and memory. This domain only accounts for 16.7% of the maximum total score for the MoCA [15]. In addition to learning and memory, the neurocognitive domain of executive functions as well as visuo-spatial skills also serve as an important indicator for individuals who have mNCD [71]. However, our results only partly confirmed construct validity for the Qmci subtests measuring these neurocognitive domains (i.e., 'verbal fluency' and 'clock drawing'). This finding is surprising, as the clock drawing test is the third most frequently cited screening test after the MoCA and MMSE [12]. However, although it has shown good reliability, validity is only fair to good [12], which aligns with our findings. In addition, the subtest 'clock drawing' was found least accurate (AUC=0.57) in discriminating HOA from mNCD [59], suggesting that there exists potential for enhancing Qmci's diagnostic accuracy by substituting the corresponding subtests with alternatives that exhibit better construct validity and discriminatory power. Similarly, the 'verbal fluency' subtest of the Qmci did not meet our criteria for verifying construct validity to measure executive functions, although these have been demonstrated to have predictive value in detecting mNCD and differentiating it from HOA [71]. Nonetheless, this subtest has been shown to be the second most accurate in discriminating HOA from mNCD with an AUC of 0.77 [59].

Although our results suggest that the Qmci subscores measuring learning and memory could be analyzed separately, it must be emphasized that this was not originally intended. Rather, the total Qmci score, which has demonstrated construct validity (as described above), should primarily be analyzed [21]. Future research should directly compare the Qmci-G total score to reference assessments to verify construct validity of the Qmci-G. In addition, future research should explore whether the Qmci's diagnostic accuracy can be further enhanced by substituting some of the subtests with an alternative that exhibits better construct validity and discriminatory power.

Generalizability of the findings

Compared to the extensive validation studies of the original English Qmci [24], our study populations were comparable regarding the descriptive statistics on age (for both HOA and mNCD), sex distribution (only for HOA), and descriptive data on the Qmci total score and subscores (only for HOA). Our population of individuals who have mNCD had a slightly lower total score compared to the English Qmci, which is explained by a lower score in the subtest 'logical memory' whereas the descriptive data on all other subtests were similar [59]. Nonetheless, we found the same optimal cut-off score compared to the English Qmci [24], whereas a large variation of optimal cut-off scores has been observed in other languages of the Qmci (Chinese=55.5 [25], Dutch=51.5 [26], Greek≤51 [28] or ≤71 [27, 28], Japanese≤61 [29], Persian≤53 [30], Taiwanese=51.5 [31], and Turkish≤53 [32]). These differences are likely related to differences in socio-demographic factors and/or the small sample sizes of these studies.

In this regard, our study population had substantially more years of education (in both groups) and men were overrepresented in the group of individuals who have mNCD. While a higher level of education is a well-known early-life protective factor against mNCD and might be linked to better cognitive performance and higher cognitive reserve [72, 73], there was no relevant between-group difference in years of education which aligns with the extensive validation studies of the original English Qmci [24]. Nonetheless, previous research has shown that the optimal cut-off values as well as sensitivity and specificity of the Qmci differ between groups of varying educational levels [24]. Additionally, overrepresentation of men may influence generalizability of our findings, because women have a higher prevalence of non-amnestic mNCD [74]. However, most of our study participants had mNCD due to Alzheimer's disease and previous research has shown that there were no significant sex differences in the prevalence or incidence of mNCD when all subtypes were combined [74]. The reason for our well-educated and men-dominated study population of individuals with mNCD might be related to a selection bias

in recruitment of study participants, because this study was conducted in a research environment and only participants that were referred to us by our clinical collaboration partners could be considered. This might limit the generalizability of our findings, especially in individuals with low levels of education and to some degree also in women.

On the other hand, the distribution of etiologies of mNCD was representative, as approximately 60–90% of individuals with mNCD have Alzheimer's disease etiology, mild vascular neurocognitive disorder is the second most common etiology of mNCD, and only about 5% have mild frontotemporal neurocognitive disorder etiology [5].

To summarize, our findings are generalizable to moderately to highly educated populations with all etiologies of mNCD and across a wide age range. However, the generalizability of our findings may be limited in less educated individuals, women, and non-research settings. Therefore, it seems of crucial importance to verify these findings for the Qmci-G by testing it in clinical environments and/or primary health care in more representative populations of individuals who have mNCD.

Implications for Research and Clinical Practice

The Qmci presents as an interesting avenue for improving early detection of individuals with suspected mNCD thanks to its shorter administration time compared to the most commonly used screening tools for suspected NCDs [8, 11–13], coupled with comparable [11] or even marginally superior [23] diagnostic accuracy. Additionally, the published guidelines for the administration and evaluation of the Qmci [21] are well-developed and allow for easy administration, scoring, and interpretation. This promotes its widespread use in research and clinical settings. However, prior to the widespread implementation of the Qmci-G in clinical practice in German-speaking countries, it is crucial to verify our findings on good diagnostic accuracy, reliability, and construct validity in clinical settings and/or primary healthcare. In this regard, it is recommended that the test-retest reliability of the Qmci-G be assessed to calculate the minimal detectable difference and ultimately determine the minimum clinically relevant change. In addition, direct comparisons with standard screening tools commonly used in these settings, such as the German version of the MoCA, are necessary to determine whether the Qmci-G outperforms these screening tools. The potential for further enhancing the Qmci's diagnostic accuracy should also be explored by substituting some of the subtests with alternatives that exhibit better construct validity and discriminatory power. Finally, the feasibility, acceptability, and effectiveness of the implementation of the Qmci in standard clinical practice should be investigated. These investigations

have the potential to facilitate early detection of individuals with suspected mNCD and their referral for clinical diagnosis [7], which supports the implementation of lifestyle changes and/or interventions to ameliorate secondary prevention of mNCD and ultimately reduce the global burden of dementia [1].

Strengths and limitations

The major strength of this study is that we only included data of study participants who have a biomarker supported characterization of the etiology of mNCD in addition to the clinical diagnosis of mNCD according to the ICD-XI [40] or DSM-5® [7]. In addition, we included data of individuals who have different etiologies of mNCD. Both strengths increase the generalizability of our findings.

The study also has some key limitations that are worth mentioning. Most importantly, we did not conduct an a-priori sample size calculation, and the sample size was comparatively small. Such limitations may affect the robustness and generalizability of our findings, considering that the study was not designed to ensure adequate statistical power; however, our primary objective was to corroborate and extend previous research, and we carefully discussed the generalizability of our findings to ensure robust conclusions. Furthermore, we conducted a retrospective analysis of data obtained from three different studies, which involved diverse outcome assessors and additional assessments beyond those analyzed in this study. This could possibly introduce some bias to our findings. Nevertheless, the Qmci-G was consistently administered as one of the first three assessments, thereby mitigating potential participant fatigue-related bias. All additional evaluations were routinely conducted in the same standardized order. Additionally, we utilized consistent eligibility criteria and strictly adhered to specific working instructions to standardize all measurement procedures and participant instructions, minimizing bias during outcome measure assessment. Therefore, these limitations were not anticipated to significantly affect the findings. However, the study's design may have introduced selection bias in participant recruitment, particularly among those with mNCD, due to the heightened barriers to enrolling in a 12-week intervention study as opposed to a typical cross-sectional study utilized for evaluating the diagnostic accuracy of a screening tool. This may explain the limited generalizability of our findings in individuals with lower levels of education. Finally, the standard statistical significance threshold of $p \leq 0.05$ was utilized. To ensure a careful interpretation of the findings, we based our interpretation on predefined criteria that included effect size estimates with $CI_{95\%}$ combined with the significance level. Additionally, we calculated p-values only for differences between

sociodemographic factors and secondary outcomes (i.e., construct validity analysis). Therefore, this limitation did not affect our primary findings.

Conclusion

Our findings corroborate the existing evidence of the Qmci's good diagnostic accuracy, reliability, and construct validity. Additionally, the Qmci shows potential in resolving the limitations of commonly used screening tools, such as the MoCA. To verify these findings for the Qmci-G, testing in clinical environments and/or primary health care and direct comparisons with standard screening tools utilized in these settings are warranted.

Abbreviations

ACE-R	Addenbrooke's Cognitive Examination Revised
AUC	Area under the curve
BMI	Body mass index
$CI_{95\%}$	95% Confidence Interval
CERAD	Consortium to Establish a Registry for Alzheimer's Disease Battery
DSM-5®)	Diagnostic and Statistical Manual of Mental Disorders 5th Edition
HOA	healthy older adults
HOTAP-A	HOTAP picture-sorting test part A
ICD-XI	International Classification of Diseases 11th Revision
MMSE	Mini-Mental State Examination
mNCD	mild neurocognitive disorder
MNCD	major neurocognitive disorder
MoCA	Montreal Cognitive Assessment
NCD	neurocognitive disorder
NPV	negative predictive value
PEBL	Psychology Experiment Building Language
PEBL-DSB	computerized version of the Digit Span Backward test
PEBL-MRT	computerized Mental Rotation Task
PEBL-TMT-B	computerized version of the Trail Making Test – Part B
PPV	positive predictive value
Qmci	Quick mild cognitive impairment screen
Qmci-G	German version of the Quick mild cognitive impairment screen
r	Pearson's correlation coefficients
RCT	randomized controlled trial
ROC	Receiver operating characteristic
r_s	Spearman's rank correlation coefficients
SADAS-Cog	standardized Alzheimer's Disease Assessment Scale - cognitive subscale
WMS-IV-LM	subtest 'logical memory' of the Wechsler Memory Scale – fourth edition

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12877-024-05219-3.

Supplementary Material 1

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
All original study procedures were carried out in accordance with the Declaration of Helsinki. The original study protocols were approved by the ETH Zurich Ethics Committee (EK-2020-N-158 for our cross-sectional study and EK 2021-N-79 for our pilot RCT) as well as the Ethics Committees of Zurich and Eastern Switzerland (EK-2022-00386) for the RCT.

Consent for publication
Not applicable.

Author details
[1]Motor Control and Learning Group - Institute of Human Movement Sciences and Sport, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland
[2]Department of Health, OST - Eastern Swiss University of Applied Sciences, St. Gallen, Switzerland
[3]Division of Physiotherapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden

References
1. Shah H, Albanese E, Duggan C, Rudan I, Langa KM, Carrillo MC, et al. Research priorities to reduce the global burden of dementia by 2025. Lancet Neurol. 2016;15(12):1285–94. https://doi.org/10.1016/S1474-4422(16)30235-6.
2. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med. 2014;275(3):214–28. https://doi.org/10.1111/joim.12190.
3. Sachdev PS, Blacker D, Blazer DG, Ganguli M, Jeste DV, Paulsen JS, et al. Classifying neurocognitive disorders: the DSM-5 approach. Nat Rev Neurol. 2014;10(11):634–42. https://doi.org/10.1038/nrneurol.2014.181.
4. Sachs-Ericsson N, Blazer DG. The new DSM-5 diagnosis of mild neurocognitive disorder and its relation to research in mild cognitive impairment. Aging Ment Health. 2015;19(1):2–12. https://doi.org/10.1080/13607863.2014.920303.
5. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub; 2013.
6. World Health Organization. ICD-11 International Classification of Diseases 11th Revision The global standard for diagnostic health information. https://icd.who.int/en (2018). Accessed July 20 2020.
7. Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017;7(2):e011146. https://doi.org/10.1136/bmjopen-2016-011146.

8.   Chen Y-X, Liang N, Li X-L, Yang S-H, Wang Y-P, Shi N-N. Diagnosis and treatment for mild cognitive impairment: a systematic review of clinical practice guidelines and Consensus statements. Front Neurol. 2021;12. https://doi.org/10.3389/fneur.2021.719849.

9.   Force UPST. Screening for cognitive impairment in older adults: US Preventive Services Task Force Recommendation Statement. JAMA. 2020;323(8):757–63. https://doi.org/10.1001/jama.2020.0435.

10.   Care CTFPH, Pottie K, Rahal R, Jaramillo A, Birtwhistle R, Thombs BD, et al. Recommendations on screening for cognitive impairment in older adults. Can Med Assoc J. 2016;188(1):37–46. https://doi.org/10.1503/cmaj.141165.

11.   Breton A, Casey D, Arnaoutoglou NA. Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of diagnostic accuracy studies. Int J Geriatr Psychiatry. 2019;34(2):233–42.

12.   Chun CT, Seward K, Patterson A, Melton A, MacDonald-Wicks L. Evaluation of available cognitive tools used to measure mild cognitive decline: a scoping review. Nutrients. 2021;13(11):3974.

13.   Abd Razak M, Ahmad N, Chan Y, Kasim NM, Yusof M, Ghani MA, et al. Validity of screening tools for dementia and mild cognitive impairment among the elderly in primary health care: a systematic review. Public Health. 2019;169:84–92. https://doi.org/10.1016/j.puhe.2019.01.001.

14.   Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189–98. https://doi.org/10.1016/0022-3956(75)90026-6.

15.   Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–9.

16.   Pinto TCC, Machado L, Bulgacov TM, Rodrigues-Júnior AL, Costa MLG, Ximenes RCC, et al. Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly? Int Psychogeriatr. 2019;31(4):491–504. https://doi.org/10.1017/S1041610218001370.

17.   Ozer S, Young J, Champ C, Burke M. A systematic review of the diagnostic test accuracy of brief cognitive tests to detect amnestic mild cognitive impairment. Int J Geriatr Psychiatry. 2016;31(11):1139–50. https://doi.org/10.1002/gps.4444.

18.   Islam N, Hashem R, Gad M, Brown A, Levis B, Renoux C, et al. Accuracy of the Montreal Cognitive Assessment tool for detecting mild cognitive impairment: a systematic review and meta-analysis. Alzheimer's Dement. 2023;19(7):3235–43. https://doi.org/10.1002/alz.13040.

19.   Thomann AE, Berres M, Goettel N, Steiner LA, Monsch AU. Enhanced diagnostic accuracy for neurocognitive disorders: a revised cut-off approach for the Montreal Cognitive Assessment. Alzheimers Res Ther. 2020;12(1):39. https://doi.org/10.1186/s13195-020-00603-8.

20.   O'Caoimh R. The quick mild cognitive impairment (Qmci) screen: developing a new screening test for mild cognitive impairment and dementia. University College Cork; 2015.

21.   O'Caoimh R, Molloy DW. The Quick Mild Cognitive Impairment Screen (Qmci). Cognitive Screening Instruments. 2017. pp. 255 – 72.

22.   O'Caoimh R, Gao Y, McGlade C, Healy L, Gallagher P, Timmons S, et al. Comparison of the quick mild cognitive impairment (Qmci) screen and the SMMSE in screening for mild cognitive impairment. Age Ageing. 2012;41(5):624–9. https://doi.org/10.1093/ageing/afs059.

23.   Glynn K, Coen R, Lawlor BA. Is the quick mild cognitive impairment screen (QMCI) more accurate at detecting mild cognitive impairment than existing short cognitive screening tests? A systematic review of the current literature. Int J Geriatr Psychiatry. 2019;34(12):1739–46. https://doi.org/10.1002/gps.5201.

24.   O'Caoimh R, Gao Y, Svendovski A, Gallagher P, Eustace J, Molloy DW. Comparing approaches to optimize cut-off scores for short cognitive Screening instruments in mild cognitive impairment and dementia. J Alzheimers Disease. 2017;57(1):123–33. https://doi.org/10.3233/Jad-161204.

25.   Xu Y, Yi L, Lin Y, Peng S, Wang W, Lin W, et al. Screening for cognitive impairment after stroke: validation of the Chinese Version of the quick mild cognitive impairment screen. Front Neurol. 2021;12. https://doi.org/10.3389/fneur.2021.608188.

26.   Bunt S, O'Caoimh R, Krijnen WP, Molloy DW, van der Goodijk GP, et al. Validation of the Dutch version of the quick mild cognitive impairment screen (Qmci-D). BMC Geriatr. 2015;15(1):115. https://doi.org/10.1186/s12877-015-0113-1.

27.   Messinis L, Nasios G, Mougias A, Patrikelis P, Malefaki S, Panagiotopoulos V, et al. Comparison of the Greek Version of the Quick Mild Cognitive Impairment

28.   Messinis L, O'Donovan MR, Molloy DW, Mougias A, Nasios G, Papathanasopoulos P, et al. Comparison of the Greek version of the quick mild cognitive impairment screen and standardised mini-mental state examination. Arch Clin Neuropsychol. 2021;36(4):578–86.

29.   Morita A, O'Caoimh R, Murayama H, Molloy DW, Inoue S, Shobugawa Y, et al. Validity of the Japanese version of the quick mild cognitive impairment screen. Int J Environ Res Public Health. 2019;16(6):917.

30.   Rezaei M, Shariati B, Molloy DW, O'Caoimh R, Rashedi V. The Persian Version of the quick mild cognitive impairment screen (Q mci-Pr): Psychometric Properties among Middle-aged and older Iranian adults. Int J Environ Res Public Health. 2021;18(16):8582.

31.   Lee M-T, Chang W-Y, Jang Y. Psychometric and diagnostic properties of the Taiwan version of the quick mild cognitive impairment screen. PLoS ONE. 2018;13(12):e0207851.

32.   Yavuz BB, Varan HD, O'Caoimh R, Kizilarslanoglu MC, Kilic MK, Molloy DW, et al. Validation of the Turkish version of the quick mild cognitive impairment screen. Am J Alzheimer's Disease Other Dementias®. 2017;32(3):145–56. https://doi.org/10.1177/1533317517691122.

33.   Manser, P., de Bruin, E.D. Test-retest reliability and validity of vagally-mediated heart rate variability to monitor internal training load in older adults: a within-subjects (repeated-measures) randomized study. BMC Sports Sci Med Rehabil 16, 141 (2024). https://doi.org/10.1186/s13102-024-00929-y.

34.   Manser P, Poikonen H, de Bruin ED. Feasibility, usability, and acceptance of Brain-IT—a newly developed exergame-based training concept for the secondary prevention of mild neurocognitive disorder: a pilot randomized controlled trial. Front Aging Neurosci. 2023;15. https://doi.org/10.3389/fnagi.2023.1163388.

35.   Manser P, de Bruin ED. Brain-IT: Exergame training with biofeedback breathing in neurocognitive disorders. Alzheimer's Dement. 2024. https://doi.org/10.1002/alz.13913. ;n/a(n/a).

36.   Manser P, Michels L, Schmidt A, Barinka F, de Bruin ED. Effectiveness of an Individualized Exergame-Based Motor-Cognitive Training Concept targeted to improve cognitive functioning in older adults with mild neurocognitive disorder: study protocol for a Randomized Controlled Trial. JMIR Res Protoc. 2023;12:e41173. https://doi.org/10.2196/41173.

37.   Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ : Br Med J. 2015;351:h5527. https://doi.org/10.1136/bmj.h5527.

38.   Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6(11):e012799. https://doi.org/10.1136/bmjopen-2016-012799.

39.   Petermann F, Lepach AC. Wechsler Memory Scale® – Fourth Edition (WMS®-IV) - Manual zur Durchführung und Auswertung (Deutsche Übersetzung und Adaptation der WMS®-IV von David Wechsler). Pearson Assessment and Information GmbH; 2012.

40.   Wechsler D. Wechsler memory scale–fourth edition (WMS-IV). New York, NY: The Psychological Corporation; 2009.

41.   Mueller ST, Piper BJ. The psychology experiment Building Language (PEBL) and PEBL test battery. J Neurosci Methods. 2014;222:250–9. https://doi.org/10.1016/j.jneumeth.2013.10.024.

42.   Croschere J, Dupey L, Hilliard M, Koehn H, Mayra K. The effects of time of day and practice on cognitive abilities: Forward and backward Corsi block test and digit span. PEBL Technical Report Series. 2012.

43.   Mueller ST. PEBL: The Psychology experiment building language (Version 0.14) [Computer experiment programming language]. http://pebl.sourceforge.net. (2014). Accessed January 2020.

44.   Menzel-Begemann A. HOTAP–Handlungsorganisation und Tagesplanung. Testverfahren zur Erfassung der Planungsfähigkeit im Alltag: Göttingen. 2009.

45.   Berteau-Pavy D, Raber J, Piper B. Contributions of age, but not sex, to mental rotation performance in a community sample. sn]. Disponível em:< http://sites.google.com/site/pebltechnicalreports … ; 2011.

46.   Shepard RN, Metzler JJS. Mental rotation of three-dimensional objects. 1971;171(3972):701–3.

47.   Thompson CB. Descriptive data analysis. Air Med J. 2009;28(2):56–9. https://doi.org/10.1016/j.amj.2008.12.001.

48.   Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth. 2019;22(1):67–72. https://doi.org/10.4103/aca.ACA_157_18.

49. Field A, Miles J, Field Z. Discovering statistics using R. Sage; 2012.
50. Rosenthal R. Meta-Analytic Procedures for Social Research. Thousand Oaks, California: 1991. https://doi.org/10.4135/9781412984997.
51. Cohen J. Statistical power analysis for the behavioral sciences. Routledge; 1988.
52. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–5.
53. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F, OptimalCutpoints. An R Package for selecting Optimal cutpoints in Diagnostic tests. J Stat Softw. 2014;61(8):1–36. https://doi.org/10.18637/jss.v061.i08.
54. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77. https://doi.org/10.1186/1471-2105-12-77.
55. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery. 2016;159(6):1638–45. https://doi.org/10.1016/j.surg.2015.12.029.
56. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied logistic regression. Wiley; 2013.
57. Taherdoost H. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. Int J Acad Res Manage (IJARM). 2016. https://doi.org/10.2139/ssrn.3205040.
58. Standish TIM, Molloy DW, Cunje A, Lewis DL. Do the ABCS 135 short cognitive screen and its subtests discriminate between normal cognition, mild cognitive impairment and dementia? Int J Geriatr Psychiatry. 2007;22(3):189–94. https://doi.org/10.1002/gps.1659.
59. O'Caoimh R, Gao Y, Gallagher PF, Eustace J, McGlade C, Molloy DW. Which part of the quick mild cognitive impairment screen (Qmci) discriminates between normal cognition, mild cognitive impairment and dementia? Age Ageing. 2013;42(3):324–30. https://doi.org/10.1093/ageing/aft044.
60. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. USA: Oxford University Press; 2015.
61. O'Caoimh R, Molloy DW. The quick mild cognitive impairment screen (Qmci). Cognitive Screening Instruments: A Practical Approach. 2017.
62. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's cognitive examination revised (ACE-R): a brief cognitive test battery for dementia screening. Int J Geriatr Psychiatry. 2006;21(11):1078–85. https://doi.org/10.1002/gps.1610.
63. Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, et al. A total score for the CERAD neuropsychological battery. Neurology. 2005;65(1):102–6. https://doi.org/10.1212/01.wnl.0000167607.63000.38.
64. O'Caoimh R, Timmons S, Molloy DW. Screening for mild cognitive impairment: comparison of MCI Specific Screening instruments. J Alzheimers Disease. 2016;51(2):619–29. https://doi.org/10.3233/Jad-150881.
65. Fillenbaum GG, Mohs R. CERAD (Consortium to establish a Registry for Alzheimer's Disease) Neuropsychology Assessment Battery: 35 years and counting. J Alzheimers Dis. 2023;93(1):1–27. https://doi.org/10.3233/jad-230026.
66. Iavarone A, Carpinelli Mazzi M, Russo G, D'Anna F, Peluso S, Mazzeo P, et al. The Italian version of the quick mild cognitive impairment (Q mci-I) screen: normative study on 307 healthy subjects. Aging Clin Exp Res. 2019;31:353–60.
67. O'Caoimh R, Svendrovski A, Johnston BC, Gao Y, McGlade C, Eustace J, et al. The quick mild cognitive impairment screen correlated with the standardized Alzheimer's Disease Assessment Scale–cognitive section in clinical trials. J Clin Epidemiol. 2014;67(1):87–92. https://doi.org/10.1016/j.jclinepi.2013.07.009.
68. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am J Psychiatry. 1984;141(11):1356–64. https://doi.org/10.1176/ajp.141.11.1356.
69. Standish TIM, Molloy DW, Bédard M, Layne EC, Murray EA, Strang D. Improved reliability of the standardized Alzheimer's Disease Assessment Scale (SADAS) compared with the Alzheimer's Disease Assessment Scale (ADAS). J Am Geriatr Soc. 1996;44(6):712–6. https://doi.org/10.1111/j.1532-5415.1996.tb01838.x.
70. Morris JC. The clinical dementia rating (CDR): current version and scoring rules. Neurology. 1993;43(11):2412–4. https://doi.org/10.1212/WNL.43.11.2412-a.
71. Chehrehnegar N, Nejati V, Shati M, Rashedi V, Lotfi M, Adelirad F, et al. Early detection of cognitive disturbances in mild cognitive impairment: a systematic review of observational studies. Psychogeriatrics. 2019;20(2):212–28. https://doi.org/10.1111/psyg.12484.
72. Zhang Y-R, Xu W, Zhang W, Wang H-F, Ou Y-N, Qu Y, et al. Modifiable risk factors for incident dementia and cognitive impairment: an umbrella review of evidence. J Affect Disord. 2022;314:160–7. https://doi.org/10.1016/j.jad.2022.07.008.
73. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol. 2012;11(11):1006–12. https://doi.org/10.1016/S1474-4422(12)70191-6.
74. Au B, Dale-McGrath S, Tierney MC. Sex differences in the prevalence and incidence of mild cognitive impairment: a meta-analysis. Ageing Res Rev. 2017;35:176–99. https://doi.org/10.1016/j.arr.2016.09.005.

Publisher's Note