



Instruments to Assess Cognitive Reserve Among Older Adults: a Systematic Review of Measurement Properties

Wanrui Wei¹ · Kairong Wang¹ · Jiyuan Shi¹ · Zheng Li¹

Received: 16 June 2022 / Accepted: 27 March 2023 / Published online: 28 April 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Cognitive reserve explains the differences in the susceptibility to cognitive impairment related to brain aging, pathology, or insult. Given that cognitive reserve has important implications for the cognitive health of typically and pathologically aging older adults, research needs to identify valid and reliable instruments for measuring cognitive reserve. However, the measurement properties of current cognitive reserve instruments used in older adults have not been evaluated according to the most up-to-date **C**onsensus-based **S**tandards for the selection of health status **M**easurement **I**nstruments (COSMIN). This systematic review aimed to critically appraise, compare, and summarize the quality of the measurement properties of all existing cognitive reserve instruments for older adults. A systematic literature search was performed to identify relevant studies published up to December 2021, which was conducted by three of four researchers using 13 electronic databases and snowballing method. The COSMIN was used to assess the methodological quality of the studies and the quality of measurement properties. Out of the 11,338 retrieved studies, only seven studies that concerned five instruments were eventually included. The methodological quality of one-fourth of the included studies was doubtful and three-seventh was very good, while only four measurement properties from two instruments were supported by high-quality evidence. Overall, current studies and evidence for selecting cognitive reserve instruments suitable for older adults were insufficient. All included instruments have the potential to be recommended, while none of the identified cognitive reserve instruments for older adults appears to be generally superior to the others. Therefore, further studies are recommended to validate the measurement properties of existing cognitive reserve instruments for older adults, especially the content validity as guided by COSMIN. **Systematic Review Registration numbers:** CRD42022309399 (PROSPERO).

Keywords Cognitive reserve · Measurement properties · Psychometric properties · Surveys and questionnaires · Systematic review

Introduction

Globally, around 50 million people are currently living with dementia, and the incidence of dementia is expected to more than triple by 2050 (Patterson, 2018).

In the absence of the treatment of dementia, efforts should focus on exploring mechanisms and strategies to slow down the neuropathological changes and cognitive decline. Significantly, some older adults who met a neuropathologic diagnosis of high or intermediate likelihood of Alzheimer's dementia (AD) may perform in the cognitively normal range, based on the Braak score for neurofibrillary pathology and the Consortium to Establish a Registry for AD estimate of neuritic plaques as recommended by the National Institute on Aging/Reagan criteria (Bennett et al., 2006). Therefore, there exists a difference between the number of neuropathological changes present in the brain and the individual cognitive function (Morris et al., 1996; Negash et al., 2013).

Cognitive reserve is a property of the brain that allows for cognitive performance that is better than expected based on the degree of neuropathology, explaining the tolerance

✉ Zheng Li
zhengli@pumc.edu.cn
Wanrui Wei
weiwannuismile@126.com
Kairong Wang
wangkairong5@163.com
Jiyuan Shi
15038059737@163.com

¹ School of Nursing, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 33 Ba Da Chu Road, Shijingshan District, 100144 Beijing, China

and difference in the susceptibility to cognitive impairment related to brain aging, pathology, or insult (Stern et al., 2020). Growing empirical evidence has demonstrated that cognitive reserve and its components were associated with decreased risk of developing dementia (Almeida-Meza et al., 2021; Nelson et al., 2021; Valenzuela et al., 2011; Van Loenhoud et al., 2019; Xu et al., 2019) and slower cognitive decline in global cognitive function, episodic memory, and working memory, or visuospatial ability among older adults (Li et al., 2021; Sumowski et al., 2014). This highlights the relationship between cognitive reserve and identifying the risk of developing dementia and underlines the importance of early interventions associated with cognitive reserve to delay the onset of cognitive decline in older adults. Given that cognitive reserve has important implications in attenuating the influence of dementia-related adverse outcomes on cognition of typically and pathologically aging older adults (Arenaza-Urquijo & Vemuri, 2018, 2020), it is crucial for research to identify valid and reliable instruments for assessing cognitive reserve. However, cognitive reserve is a theoretical and hypothetical construct that is difficult to assess using direct quantitative measures, the major barrier to studying cognitive reserve lies in its assessment (Jones et al., 2011).

Cognitive reserve, in essence, is related to multiple factors. Traditionally, proxies and convenient indicators such as education, occupation, verbal IQ, and cognitive activity were often used, and these proxies have been proven to mitigate or moderate between pathology and cognitive decline (Boyle et al., 2021; Opdebeeck et al., 2016; Sajeev et al., 2016). In addition to single proxies, measurement models with composite proxies using combinations of proxies were operationalized to assess cognitive reserve as well. Nelson et al. (2021) categorized and summarized two approaches to quantify cognitive reserve, including the residual variance approach and the composite proxy approach. The former model, based on a reflective measurement model for reserve, operationalized cognitive reserve as the unexplained cognitive residual variance between predicted and actual cognitive performance considering neuropathology, which described the relationship between the latent variables (i.e., cognitive reserve) and explicit indicators. Whereas the latter model is estimated as a formative measurement model, measuring cognitive reserve by synthesizing various proxies such as years of education, IQ tests, occupation), which showed the association between the latent variable (i.e., cognitive reserve) attributed to a combination of indicators.

In addition, several standardized questionnaires based on proxy indicators were developed and validated to measure cognitive reserve specifically. Generic patient-reported outcome measures (PROMs) include the Cognitive Reserve Index questionnaire (CRIq) (Nucci et al., 2012), Cognitive Reserve Questionnaire (CRQ) (Rami et al., 2011), Cognitive

Reserve Scale (CRS) (León et al., 2011), Cognitive Reserve Assessment Scale in Health (CRASH) (Amoretti et al., 2019), Lifetime of Experiences Questionnaire (LEQ) (Valenzuela & Sachdev, 2007), Premorbid Cognitive Abilities Scale (PCAS) (Apolinario et al., 2013) and Retrospective Indigenous Childhood Enrichment (RICE) (Minogue et al., 2018). These proxy indicators in the PROMs are formative and lead to the development of cognitive reserve (Stern et al., 2020), and the causal ordering held that accumulation of these proxy indicators created higher cognitive reserve in late life (Jones et al., 2011). Thus, the questionnaires were based on formative measurement models, while the formative measurement model shows the association between different latent variables that consist of a combination of indicators, and the direction of causality is from items to construct (Coltman et al., 2008; Jarvis et al., 2003).

Relatively, although the residual variance approach and composite proxy approach both exert a strong effect on cognitive decline (Nelson et al., 2021), AD biomarker (e.g., hippocampal volume, gray matter volume, CSF total tau, and A β -42) and cognitive domains (e.g., attention, verbal memory, global cognitive performance and an average of standardized domain-specific tests) quantified differently in residual variance approach, additionally, proxies qualitatively varied in the composite proxy approach including measures such as vocabulary subtest of the Wechsler Adult Intelligence Scale-Revised, American National Adult Reading Test, occupation level, and years of education. Thus, they may often result in inconsistency and heterogeneity, making it difficult to compare different research findings. Since heterogeneity in cognitive reserve measures could indicate that some of them are not valid or reliable, efforts need to explore standardized methods to better measure cognitive reserve in ethnically and racially diverse older adults, which might make up for the shortcoming of the lack of uniform cognitive reserve measurement methods. Hence, it is helpful to use standardized measurement to assess cognitive reserve. Furthermore, COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) provides the methodology to assess the methodological quality of the included studies and measurement properties, which could help the selection of instruments for a given situation in clinical practice and research.

To our knowledge, only three systematic reviews on cognitive reserve instruments have been published (Kartschmit et al., 2019; Landenberger et al., 2019; Nogueira et al., 2022) summarized five instruments in the form of scales and questionnaires used to measure cognitive reserve, including CRS (León et al., 2011), CRIq (Nucci et al., 2012), CRQ (Rami et al., 2011), LEQ (Valenzuela & Sachdev, 2007) and Lifetime Cognitive Activity Scale (LCAS) (Wilson et al., 2003), but not covering all the existed cognitive reserve scales or questionnaires. What's more, Nogueira et al. (2022) focused

on the frequency of the assessments and identified the quantitative assessments of cognitive reserve, which means cognitive reserve questionnaires and scales (for example, CRIQ, CRQ, LEQ) and cognitive tests (for example, National Adult Reading Test, vocabulary scores of Wechsler Adult Intelligence Scale-Revised) that could generate a total score corresponding to the individual level of cognitive reserve. However, Landenberger et al. (2019) and Nogueira et al. (2022) differed on the instruments and did not report and analyze the psychometric properties used for the PROMs, making it difficult for clinicians and researchers to identify the optimal instrument to assess cognitive reserve. In contrast, Kartschmit et al. (2019) evaluated the psychometric properties of cognitive reserve questionnaires, nonetheless, they stated that they evaluated studies published up to 2018 using the new COSMIN checklist but rated the measurement properties of each questionnaire as ‘excellent’, ‘good’, ‘fair’, or ‘poor’, which was following the original COSMIN checklist. The methodology they used contradicted the new COSMIN as they stated, while Mokkink et al. (2018) recommended the rating score of ‘very good’, ‘adequate’, ‘doubtful’, or ‘inadequate’ quality for each of the items in the updated COSMIN checklist, because the older rating labels do not accurately reflect the judgments given, and do not match the descriptions used in the new COSMIN checklist. Thus, there is a need to rate the measurement properties following the updated COSMIN methodology rigorously and give some recommendations on the measurement of cognitive reserve in older adults.

At this point, this systematic review aimed to identify all cognitive reserve instruments which have been developed, adapted, or validated among older adults in the literature and their psychometric properties. Additionally, the methodological quality of the studies on measurement properties was also evaluated as indicated by the most up-to-date COSMIN guidelines.

Methods

This systematic review was conducted according to the COSMIN guidelines (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018), which were used as methodological guidelines for this systematic review. Three consecutive parts (i.e., Part A, Part B, and Part C) with the ten-step procedure are recommended as COSMIN methodology (Prinsen et al., 2018). Firstly, Part A involves performing the literature search, which includes the first four steps, there are formulating the aim of the review and eligibility criteria, performing the literature search, and selecting relevant articles. Secondly, in Part B, three sub-steps were used, consisting of evaluating the methodological quality of content validity, internal structure, and the remaining measurement

properties by the COSMIN Risk of Bias Checklist (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018), applying criteria for good measurement properties by using the updated criteria for good measurement properties (Prinsen et al., 2018), and summarizing the evidence and grading the evidence quality by using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach (Schünemann et al., 2013). Thirdly, Part C consists of the remaining steps, including the evaluation of the interpretability and feasibility of the PROMs, and formulating and reporting of the systematic review.

Additionally, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Page et al., 2021) as reporting guidelines for this systematic review (Supplementary File 1). This systematic review was registered on the prospective international register of systematic reviews (PROSPERO, CRD42022309399-<https://www.crd.york.ac.uk/prospero/>).

Search Strategy

A comprehensive three-step search of published studies was undertaken. An initial search was first conducted in PubMed to capture keywords by analyzing the title and abstract and index terms. Based on the findings of this search, structured searches were conducted in China National Knowledge Infrastructure (CNKI), Wanfang Data, China Science and Technology Journal Database, China Biology Medicine disc (CBMdisc), PubMed (MEDLINE), Embase (Ovid), Web of Science (Thomson Scientific), Scopus (Elsevier), Cumulative Index to Nursing and Allied Health Literature Plus (CINAHL Plus) (EBSOCO), PsycInfo (EBSOCO), Cochrane Library (Wiley), ProQuest Dissertations and Theses and medRxiv (<https://www.medrxiv.org/>).

All studies published up to 22 December 2021 were included. We developed search strategies using Medical Subject Headings terms and free text words accordingly for each database. Search terms usually involve four key elements of the review aim: (1) construct, (2) population(s), (3) type of instrument(s), and (4) measurement properties. The search terms for the population were not used to broaden the potential hits, thus, the construct (i.e., cognitive reserve) was combined with the type of instruments and measurement properties to build the search strategy and ensure that all relevant literature was included. We developed a comprehensive filter provided for measurement properties (Terwee et al., 2009) in PubMed. These search terms for cognitive reserve, type of instruments, and measurement properties were searched respectively, and then they were connected by the Boolean operator. A summary of the search strategy used in this study is provided in Supplementary File 2 in detail. Finally, we manually searched the reference list of the retrieved articles to identify additional relevant articles.

Eligibility Criteria

All original studies conducted in any country published in either English or Chinese were eligible for inclusion in this study irrespective of the sample size. The following inclusion criteria were used: (1) Type of instruments: original studies aimed to measure the cognitive reserve; (2) Type of participants: the study sample or at least 50% of them should represent the population aged 60 or above, which means a median participant age of 60 or above; (3) Type of studies: studies that evaluated at least one of the COSMIN psychometric measurement properties, including the development of a PROM (to rate the content validity), or criteria necessary for the evaluation and the interpretability of the PROMs of interest, such as the distribution of scores in the study population, the percentage of missing items, floor and ceiling effects, the availability of scores and change scores for relevant groups or subgroups, and the minimal important change or minimal important difference (Griffiths et al., 2015), were included in this systematic review. The following exclusion criteria were used: (1) Study types were conference proceedings, editorials, case reports, commentaries, reviews, abstracts, newsletters, research protocols, editorial letters; (2) Studies that provided indirect evidence of the measurement properties, or used the instruments only to measure outcomes (e.g., in randomized controlled trials), or used the PROM to validate another instrument; (3) Studies that did not include the full text.

Study Selection

EndNote 20 was used to manage the references. The abstracts and full-text articles were independently and respectively selected by two of the three reviewers (WW and KW, or WW and JS), and any disagreements were resolved through discussion between the two paired reviewers. A fourth reviewer (ZL) was consulted if a consensus was still not reached following discussions.

Data Extraction and Data Analysis

Two of the three reviewers (WW and KW, or WW and JS) independently and respectively extracted data from the included papers and appraised the studies. Any discrepancies were resolved through a discussion between the two reviewers. If the disagreement persisted, a fourth review (ZL) was consulted.

For all included studies, the characteristics of the included PROM and the results of the study on measurement properties were extracted and collected in two standardized information forms. The data collected regarding the characteristics of the included PROMs included item generation,

target population, participants (sample size), number of older adults, mode of administration (self-report, interview-based, parent or proxy report, etc.), number of items, completion time, response options, range of scores and the original language of the study. The data extracted from the results of studies on measurement properties included content validity, structural validity, internal consistency, reliability, and hypothesis testing for construct validity.

The studies were rated using the three sub-step procedures indicated by the COSMIN as described below.

Sub-Step 1: Evaluating the Methodological Quality of the Included Studies

The COSMIN Risk of Bias Checklist (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018) was used to evaluate the methodological quality of the included studies. The checklist is based on the following ten boxes: instrument development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness. The studies were scored using a four-point rating scale, including ‘very good’, ‘adequate’, ‘doubtful’, or ‘inadequate’ quality for each item in the boxes indicated by the COSMIN checklist. An overall score was obtained based on the lowest rating within the criterion called ‘the worst score counts’. Among the measurement properties, content validity is the most important measurement property of a PROM, because it refers to the extent to which the content of the instrument adequately reflects cognitive reserve, which shows relevance, comprehensiveness, and comprehensibility of the instrument.

Sub-Step 2: Applying Criteria for Good Measurement Properties

Consensus-based definitions of domains, measurement properties, and aspects of measurement properties in the COSMIN checklist are reached by the COSMIN panel (Mokkink et al., 2010). In terms of criteria for good measurement properties, the result of the single study was rated against the updated criteria for good measurement properties (Prinsen et al., 2018). The measurement properties were rated as sufficient (+), insufficient (−), or indeterminate (?). Definitions and quality criteria for measurement properties cited from COSMIN manual (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018) for systematic reviews of PROMs were presented in Supplementary File 3.

However, the COSMIN can only assess the structural validity criteria associated with confirmatory factor analysis or item response theory (IRT) and does not provide the criteria necessary for exploratory factor analysis (EFA). As COSMIN stated that additional criteria could be used for

assessing the results of studies, therefore, additional criteria were used to rate the EFA (Lee et al., 2020), including the studies were rated as sufficient (+) if the provided factors explained no less than 50% of the variance, indeterminate (?) if the explained variance was not reported, and insufficient (–) if the factors explained less than 50% of the variance. In addition, for reliability, the criteria on COSMIN (Prinsen et al., 2018) involved the intraclass correlation coefficient (ICC) and weighted Kappa but did not mention the criteria on Pearson's coefficient. Thus, Lee et al.'s criteria (2020) were referred to, including Pearson's r greater or equal to 0.80 indicating sufficient (+), and Pearson's r below 0.80 was rated as insufficient (–). If Pearson's coefficient was not provided, the study was rated as indeterminate (?). Since there were no quality criteria available to assess feasibility and interpretability, we described the data in the text.

Sub-Step 3: Summarizing the Evidence and Grading the Evidence Quality

Since sub-step 1 and sub-step 2 focused on the methodological quality of each study and measurement properties of the PROM, sub-step 3 shed light on the quality of the PROM in its entirety. At this step, all results per measurement property of a PROM are quantitatively pooled when the studies are enough or qualitatively summarized when the studies are limited.

To conclude the quality of the PROM, the consistency of each measurement property rating between each study should be considered. If the ratings were consistent between studies, the results could be pooled and an overall rating of sufficient or insufficient could be assigned to that measurement property. If the ratings were inconsistent between studies, further subgroup analysis was performed to evaluate the factors leading to inconsistencies (e.g., different languages of a PROM); however, if no reasonable explanation was found to explain the inconsistency, the overall rating of this measurement property was rated as inconsistent (\pm). On the other hand, if there was no information supporting the rating, an indeterminate (?) overall rating was assigned.

Subsequently, the evidence was summarized and graded as high, medium, low, or very low according to the modified GRADE approach (Schünemann et al., 2013). The quality of the evidence was graded for each measurement property and each PROM separately. Four of the five GRADE factors were adopted in the COSMIN methodology, including (1) Risk of bias (i.e., the methodological quality of the studies); (2) Inconsistency (i.e., unexplained inconsistency of results across studies); (3) Imprecision (i.e., the total sample size of the available studies), (4) Indirectness (i.e., evidence from different populations than the population of interest in the review). It is important to note that the imprecision principle should not be used to assess content validity, structural

validity, and cross-cultural validity because a sample size requirement has already been included in the COSMIN Risk of Bias box for the measurement properties.

In addition, interpretability is the qualitative meaning of a PROM's quantitative scores or change scores (Mokkink et al., 2010), while feasibility is the ease of using a PROM given constraints such as completion time, ease of administration, length of the instrument. Since interpretability and feasibility are not measurement properties and no quality criteria for them, we presented and described them in the text but were not evaluated as they are important aspects of choosing suitable instruments (Prinsen et al., 2018).

COSMIN recommendations for the most appropriate PROM, consisting of three categories, are based on the cognitive reserve of older adults (Prinsen et al., 2018). PROMs are rated as 'A', which is trustworthy if they have sufficient content validity at any quality level and sufficient internal consistency with evidence that is at least low-quality. PROMs are categorized as 'C' if they have high-quality evidence with an insufficient measurement property, and these PROMs are not recommended for clinical use. All other PROMs that do not fit either 'A' or 'C' would be categorized as 'B', and these PROMs have the potential to be recommended for use, requiring further validation research to assess the quality.

Results

Studies Identification

The details of the systematic search and study selection process are presented in a flow diagram (Fig. 1). The initial search strategy yielded 11,338 records. After removing duplicates, 5,795 reports were selected for screening at the title and abstract levels. A total of 5,753 studies were excluded as they did not meet the eligibility criteria of this study. After evaluating the remaining 40 full-text studies, four studies (Choi et al., 2016; Leon-Estrada et al., 2017; León et al., 2011; Rami et al., 2011) involved four instruments measuring the cognitive reserve in non-Chinese or non-English languages were identified and were therefore excluded. Moreover, after screening the full texts of the included reports, nine studies (Altieri et al., 2018; Amoretti et al., 2019; Apolinario et al., 2013; Çebi & Kulce, 2021; Kaur et al., 2021; León et al., 2014; Maiovis et al., 2016; Nucci et al., 2012; Ozakbas et al., 2021) were excluded because they did not follow the age criteria that at least 50% of the study sample older than 60 years old. An additional study (Ourry et al., 2021) was included after snowballing. Finally, seven studies were included for methodological quality assessment.

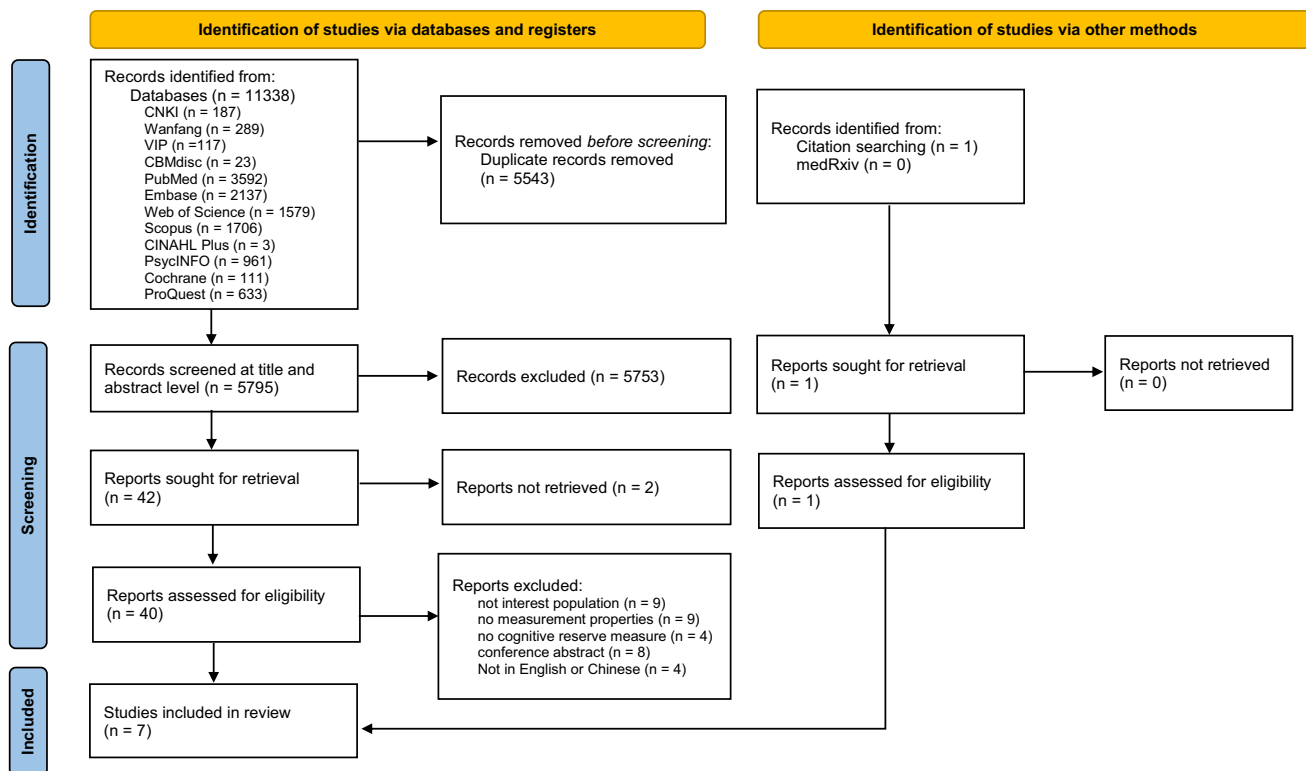


Fig. 1 PRISMA flow diagram of systematic search and study selection

Characteristics of the Included Studies

Within the seven studies, five different PROMs evaluating the cognitive reserve for older adults were identified, including the English version of CRIq (E-CRIq) (Garba et al., 2020), the modified version of CRS (mCRS) (Relander et al., 2021), LEQ (Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007), the American version of LEQ (A-LEQ) (Gonzales, 2012), and RICE (Minogue et al., 2018).

The characteristics of the included studies are summarized in Table 1. All studies were published in English between 2007 and 2021. Two of the included studies were conducted in America ($n = 2$) (Garba et al., 2020; Gonzales, 2012), and the rest were conducted in Finland ($n = 1$) (Relander et al., 2021), Australia ($n = 2$) (Minogue et al., 2018; Valenzuela & Sachdev, 2007), France, Germany, Spain and United Kingdom ($n = 1$) (Ourry et al., 2021), and India ($n = 1$) (Paplikar et al., 2020). The items used to generate the instruments were based on existing instruments, including literature reviews and interviews. These instruments were all self-reported (Garba et al., 2020; Gonzales, 2012; Minogue et al., 2018; Ourry et al., 2021; Paplikar et al., 2020; Relander et al., 2021; Valenzuela & Sachdev, 2007). The target populations included older adults attending hospital clinics or residing

in nursing homes (Garba et al., 2020), receiving inpatient care (Relander et al., 2021), and residing in either inpatient or community dwellings (Minogue et al., 2018; Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007), or independent residential facilities, senior centers or wellness centers (Gonzales, 2012). Among them, one study included people below the age of 60, however, since the median age of the participants in this study was 62 years, it was still included in the analysis (Relander et al., 2021). The studies involved older adults who were cognitively normal (Gonzales, 2012; Minogue et al., 2018; Ourry et al., 2021; Relander et al., 2021; Valenzuela & Sachdev, 2007), cognitively impaired (Garba et al., 2020), or both of them (Paplikar et al., 2020).

The instruments were composed of 18 to 42 items, generally divided into diverse domains. Three instruments (Gonzales, 2012; Ourry et al., 2021; Paplikar et al., 2020) were based on the translated version of the LEQ with 42 items (Valenzuela & Sachdev, 2007), one instrument was based on the modified LEQ instrument (Gonzales, 2012) with 36 items. The completion time of the cognitive reserve instruments ranged from 15 to 35 min. The item-response type included choices with frequency, Likert scales with five points, free responses, or a mixed combination of the three-item response types.

Table 1 Characteristics of the included studies

PROM ^a	Author, year, and country	Item generation	Target population	Participants (sample size)	Number of older adults	Mode of administration	Number of items, (sub scale(s) (number of items)	Completion time (minutes)	Response options	Range of scores	Original language
English version of Cognitive Reserve Index Questionnaire (E-CRIq)	Garba et al., 2020, America	based on the original scale (the CRIq)	Alzheimer's disease or Unspecified Dementia without Behavioral Disturbance patients	Alzheimer's disease or Unspecified Dementia without Behavioral Disturbance patients in a clinic and 3 nursing homes (n = 90)	90	interview-based, self-report	some demographic data (date and place of birth, gender, place of residence, nationality, marital status), and 20 items divided into 3 subscales (education, working activity, and leisure time)	15	the frequency (two choices: never/rare, often/always), and the number of years (how long each activity had been carried out)	Higher scores indicate greater cognitive reserve.	Italian
modified version of the Cognitive Reserve Scale (mCRS)	Relander et al., 2021, Finland	based on the original scale (the CRS)	healthy adults	neurologically healthy adults in hospital (n = 69)	36–64 years old (n = 30; 43.48%) ≥65 years old (n = 30; 43.48%)	self-report	20 items divided into 3 subscales (studying and information seeking, hobbies, and social relationships) averaged across 3 age groups [the young adulthood (18–35 years) and, if applicable, adulthood (36–64 years) and late adulthood (≥ 65 years)]	not reported	a 5-point Likert scale: 0 (never), 1 (once or several times per year), 2 (once or several times per month), 3 (1–2 times per week), 4 (almost every day)	0–80 Higher scores indicate greater cognitive reserve.	Spanish

Table 1 (continued)

PROM ^a	Author, year, and country	Item generation	Target population	Participants (sample size)	Number of older adults	Mode of administration	Number of items, (sub scale(s) (number of items)	Completion time (minutes)	Response options	Range of scores	Original language
Lifetime of Experiences Questionnaire (LEQ)	Valenzuela & Sachdev, 2007, Australia	based on the literature review	healthy older individuals	community-dwelling older adults (n = 79)	79	self-report	42 items divided into 3 subscales (young adulthood, midlife, and late-life) and 2 specificities (life-stages and specific versus non-specific mental activity) under each subscale	30	5-point Likert scale and free responses (free responses were recoded into ordinal scale scores) about the frequency and intensity of participation in each activity	Higher scores indicate greater cognitive reserve.	English
France, Germany, Spain, and United Kingdom version of Lifetime Experiences of Questionnaire (FGSU-LEQ)	Curry et al., 2021, France, Germany, Spain and United Kingdom	based on the original scale (the LEQ)	cognitively unimpaired older adults	cognitively unimpaired older adults in the community or from medical facilities (e.g., memory clinics) from 5 European centers (n = 359) ^c	359	self-report	42 items divided into 3 subscales (young adulthood, midlife, and late-life) and 2 specificities (life-stages and specific versus non-specific mental activity) under each subscale	30	5-point Likert scale and free responses (free responses were recoded into ordinal scale scores) about the frequency and intensity of participation in each activity	Higher scores indicate greater cognitive reserve.	English

Table 1 (continued)

PROM ^a	Author, year, and country	Item generation	Target population	Participants (sample size)	Number of older adults	Mode of administration	Number of items, (sub scale(s) (number of items)	Completion time (minutes)	Response options	Range of scores	Original language
Indian version of Lifetime of Experiences Questionnaire (I-LEQ)	Paplikar et al., 2020, India	based on the original scale (the LEQ)	community-dwelling older adults	community-dwelling healthy elderly (n = 52) and dementia elderly (n = 30)	82	self-report	42 items divided into 3 subscales (young adulthood, midlife, and late-life) and 2 specificities (life-stages and specific versus non-specific mental activity) under each subscale	30-35	5-point Likert scale and free responses (free responses were recoded into ordinal scale scores) about the frequency and intensity of participation in each activity	Higher scores indicate greater cognitive reserve.	English
American version of Lifetime Experience Questionnaire (A-LEQ)	Gonzales, 2012 ^b , America	based on the original scale (the LEQ)	older adults	older adults in independent residential facilities, senior centers, or wellness centers (n = 90)	90	self-report	36 items divided into 3 subscales (young adulthood, midlife, and late-life) and 2 specificities (life-stages and specific versus non-specific mental activity) under each subscale	30	5-point Likert scale and free responses (free responses were recoded into ordinal scale scores) about the frequency and intensity of participation in each activity	Higher scores indicate greater cognitive reserve.	English

Table 1 (continued)

PROM ^a	Author, year, and country	Item generation	Target population	Participants (sample size)	Number of older adults	Mode of administration	Number of items, (sub scale(s) (number of items)	Completion time (minutes)	Response options	Range of scores	Original language
Retrospective Indigenous Childhood Enrichment scale (RICE)	Minogue et al., 2018, Australia	based on the literature review and a pilot cohort of older Aboriginal people and other community members in urban and regional settings	Aboriginal Australian people	community-dwelling Aboriginal Australian people without dementia or mild cognitive impairment (study 1: n = 294, study 2: n = 30)	study 1: 294 study 2: 30	self-report	18 items divided into 3 subscales: traditional, intellectual, community	not reported	A 5-point Likert scale: 1 (never/none), 2 (rarely/few), 3 (sometimes/some), 4 (often/many), 5 (very often/a lot)	18-90 Higher scores indicate greater cognitive reserve.	English

^aEach version of a patient-reported outcome measure (PROM) is considered as a separate PROM in this table

^bDoctoral dissertation

^c46.2% of whom met criteria for subjective cognitive decline

Sub-Step 1: Results of Methodological Quality of the Included Studies

The methodologic quality of each questionnaire and measurement properties of the included studies is summarized in Table 2. None of the five instruments identified could be assessed against all measurement properties. Internal consistency, reliability, and hypotheses testing for construct validity (mainly convergent validity) were the three most frequently reported measurement properties in the included studies. No assessment of measurement error or criterion validity was made in any of the studies. Thirty measurement properties were evaluated for quality in these studies, and most of the properties were rated as ‘very good’ or ‘doubtful’. The details of the assessed instruments are outlined separately below.

The E-CRIQ was assessed in one study in which reliability and construct validity were evaluated (Garba et al., 2020). Reliability was scored as ‘adequate’ because the Pearson correlation coefficient was calculated with evidence that there was no systematic change in the variance between time points. In this study, concurrent validity was measured by the Wechsler Test of Adult Reading (WTAR), while the intelligence quotient (IQ) was used as the gold standard for cognitive reserve in Garba et al.’s study. However, IQ was not a reliable gold standard of cognitive reserve. According to the COSMIN methodology, if the term criterion validity for studies were used to compare with an instrument measuring a similar construct, in most cases, this would be considered evidence for construct validity rather than criterion validity. Thus, we assessed the convergent validity, which was deemed as ‘very good’.

The mCRS was evaluated in one study (Relander et al., 2021). Internal consistency and construct validity were scored as ‘very good’. Construct validity was based on occupation and education, which were defined as proxy indicators of cognitive reserve. Thus, this was regarded as convergent validity.

The LEQ was assessed in three studies (Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007). Internal consistency was scored highly in the Valenzuela and Sachdev (2007) and Paplikar et al. (2020) studies, while responsiveness was evaluated as ‘very good’ in the study of Valenzuela and Sachdev (2007) as well. The PROM development of Valenzuela and Sachdev (2007) was rated as ‘inadequate’ because there was no cognitive interview study or other pilot tests to evaluate the comprehensibility and comprehensiveness of LEQ. On the other hand, the PROM in the study of Paplikar et al. (2020) was rated as ‘doubtful’, because the comprehensibility and comprehensiveness of the PROM were not evaluated despite a pilot study performed to ensure its applicability among individuals with diverse cultural and socioeconomic backgrounds.

Table 2 Methodological quality assessment of the included studies per measurement property per patient-reported outcome measure

Questionnaire, author, and year	PROM development	Content validity	Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
E-CRIq										
Garba et al. (2020)					A				V	
mCRS										
Relander et al. (2021)				V					V	
LEQ										
Valenzuela and Sachdev (2007)	I	I	D	V	D				V	V
Ourry et al. (2021)			A		I				V	
Paplikar et al. (2020)	D	D		V	D				V	
A-LEQ										
Gonzales (2012)			I	V	D				V	
RICE										
Minoque et al. (2018)	D	I	A	V	I	A			V	

Abbreviations: V = very good; A = adequate; D = doubtful; I = inadequate; E-CRIq = English version of Cognitive Reserve Index Questionnaire; mCRS = modified version of the Cognitive Reserve Scale; LEQ = Lifetime of Experiences Questionnaire; A-LEQ = American version of Lifetime Experience Questionnaire; RICE = Retrospective Indigenous Childhood Enrichment scale

In terms of content validity, the study by Valenzuela and Sachdev (2007) was rated as ‘inadequate’ as no information was provided on how the content validity was performed. Conversely, the content validity of the Paplikar et al. (2020) study was rated as ‘doubtful’ as they reported the face validity instead. Additionally, the face validity was measured by an expert committee using a subjective assessment tool that did not specify the exact items used for the evaluation. The structural validity of the study by Valenzuela and Sachdev (2007) was scored as ‘doubtful’ because the sample size ($n = 79$) was lower than seven times the number of items and the sample size of 100 was not appropriate, while the study by Ourry et al. (2021) used Dual Multiple Factor Analysis that is exploratory factor analysis, so the score methodological quality of structural validity was rated as ‘adequate’. The cross-cultural validity of Ourry et al. (2021) was ‘inadequate’ because the sample size included in the analysis was less than 100 subjects per group. The reliability of the Valenzuela and Sachdev (2007) and Paplikar et al. (2020) studies were judged as ‘doubtful’ since the appropriateness of the time interval on the test-retest design was not stated, the COSMIN Risk of Bias Checklist standardized that an appropriate time interval should get rid of recall bias and ensure that the cognitive reserve of older adults has not altered.

The A-LEQ (Gonzales, 2012) was evaluated in one study, in which four measurement properties were scored. The construct validity was based on the comparison with the Mini-Mental State Exam (MMSE). However, Gonzales (2012) stated MMSE was an instrument used to assess convergent validity; thus, we assessed the association as convergent validity. The methodological quality of the structural validity was ‘inadequate’ because of the limited sample size ($n = 90$). The reliability was ‘doubtful’ as Pearson’s correlation coefficient was used to assess the retest reliability of the LEQ.

The RICE was evaluated in one study (Minogue et al., 2018), in which seven measurement properties were scored. The internal consistency and construct validity were deemed as ‘very good’. The PROM development was rated as ‘doubtful’ because the details of the pilot study were not reported. The relevance and comprehensiveness for professionals were not assessed for content validity, so the score in the COSMIN checklist was ‘inadequate’. The structural validity was ‘adequate’ because EFA was performed. The measurement invariance based on the different modes of administration was scored as ‘inadequate’ because the sample size was smaller than 100 per group.

Sub-Steps 2 and 3: Results of the Measurement Properties and Quality of Evidence of the PROMs

The assessment of the measurement properties and the quality of evidence are presented in Table 3. Due to limited

Table 3 Summary of findings and quality of evidence per measurement property per patient-reported outcome measure

Questionnaire, author, and year	Content validity	Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
E-CRIq									
Garba et al. (2020)					+			-	
Overall rating					+			-	
Quality of evidence					Low			Moderate	
mCRS									
Relander et al. (2021)			?					+	
Overall rating			?					+	
Quality of evidence			Low					Low	
LEQ									
Valenzuela and Sachdev (2007)	+	?	?		+			+	+
Urury et al. (2021)		+		+				+	
Paplikar et al. (2020)	+		?		+			+	
Overall rating	+	±	?	+	+			+	+
Quality of evidence	Low	Low	High	Very low	Moderate			High	Moderate
A-LEQ									
Gonzales (2012)		+	?		+			+	
Overall rating		+	?		+			+	
Quality of evidence		Very low	Moderate		Very low			Moderate	
RICE									
Minogue et al. (2018)	+	?	?	+	+			+	
Overall rating	+	?	?	+	+			+	
Quality of evidence	Very low	Moderate	High	Very low	Moderate			High	

Abbreviations: + = Sufficient; ? = insufficient; ± = indeterminate; - = insufficient; ? = inconsistent; ± = inconsistent; E-CRIq = English version of Cognitive Reserve Index Questionnaire; mCRS = modified version of the Cognitive Reserve Scale; LEQ = Lifetime of Experiences Questionnaire; A-LEQ = American version of Lifetime Experience Questionnaire; RICE = Retrospective Indigenous Childhood Enrichment scale

studies using PROMs to measure cognitive reserve in older adults, it was impossible to perform statistical analysis on the pooled results. Therefore, instead, we reported the measurement properties and the best evidence synthesis per measurement property separately. Furthermore, there was no formal information on the interpretability of the cognitive reserve instruments in the studies. The completion time, length of the instrument, type of administration, and availability in different settings were summarized regarding the feasibility (Table 1). However, no information was provided concerning the patients' or clinicians' comprehensibility and the ease of standardization, calculation, and administration of each instrument.

The E-CRIq is an English version instrument used to assess cognitive reserve in people with AD. The instrument scored low evidence for sufficient reliability and moderate evidence for insufficient construct validity (Garba et al., 2020). The study showed an insignificant, weak positive correlation statistically ($r=0.30$, $p=0.10$) between scores obtained using the WTAR and E-CRIq.

The mCRS is a 20-item scale without the items used to assess daily living in the original instrument. The quality of the evidence for this instrument was rated as low for structural validity and construct validity (Relander et al., 2021). Moreover, the quality of evidence for structural validity and construct validity was rated as low, as it included young adults and due to imprecision caused by the very small sample size ($n < 100$). Furthermore, in the absence of structural validity, the quality of the internal consistency remained indeterminate.

The LEQ (Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007) was developed to measure the cognitive reserve of older adults. The three studies that used the LEQ were conducted in Australia, France, Germany, Spain, the United Kingdom, and India. Given the risk of bias in the methodological quality of studies, the quality of the evidence regarding sufficient content validity was low. The structural validity with low-quality evidence was inconsistent because Valenzuela and Sachdev (2007) did not report the model fit for IRT. Therefore, internal consistency was rated as indeterminate, as this study did not meet the criteria of at least low evidence for sufficient structural validity. The cross-cultural validity for the LEQ was rated as sufficient with very low evidence quality, as the small sample size per group led to an extremely serious risk of bias.

The A-LEQ was modified from the LEQ, of which the number of items was reduced from 42 to 36 (Gonzales, 2012). The structural validity score was 'inadequate' for the A-LEQ, and therefore the evidence quality of this instrument was rated as very low. The internal consistency rating remained indeterminate without at least low-quality structural validity. Due to the sample size of the study being below 100, the imprecision rating was downgraded by one level, thus, the

internal consistency and hypotheses testing for the construct validity showed a moderate quality of evidence. Additionally, the study had sufficient reliability but was rated as having a very low quality of evidence due to the high risk of bias and imprecision caused by the small sample size.

RICE was designed for the Australian aboriginal population (Minogue et al., 2018). The content validity for this instrument was rated as sufficient, and the evidence quality was rated as very low due to the very serious risk of bias. The measurement invariance and hypotheses testing for construct validity were rated as high quality with sufficient ratings.

Discussion

To the best of our knowledge, this is the first systematic review aimed at evaluating the methodological quality of studies and the rating of measurement properties on instruments for cognitive reserve in older adults based on the most up-to-date COSMIN methodology (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). This systematic literature review identified five instruments to assess cognitive reserve in older adults. Nevertheless, their quality should undergo further validation.

Quality of Measurement Properties

The most frequently evaluated evidence quality properties were structural validity, internal consistency, reliability, and hypotheses testing for construct validity. However, more than half of the measurement properties were not assessed, and none of the studies assessed the measurement error and criterion validity. In addition, the methodological quality of the three-seventh of the included studies was 'very good', while the methodological quality of the studies was rated as 'doubtful' for a quarter of the measurement properties. The quality of the evidence was rated as low or very low for about half of the measurement properties and as high in only four of the measurement properties.

Content validity on cognitive reserve measurement was scarce in older adults, as none of the PROMs provided high- or moderate-quality evidence for content validity in this systematic review. E-CRIq (Garba et al., 2020), mCRS (Relander et al., 2021), and A-LEQ (Gonzales, 2012) were modified or adapted versions of the original one without content validity, only LEQ (Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007) and RICE (Minogue et al., 2018) provided low evidence of sufficient content validity. While content validity is deemed the utmost important measurement property as it shows the relevance, comprehensiveness, and comprehensibility of the instrument, this lack of direct assessment of content

validity means that the relevance or comprehensiveness or comprehensibility of all these PROMs should be further assessed on cognitive reserve among older adults.

Most of the information regarding the structural validity and internal consistency in the evaluated studies was either inconsistent or indeterminate. Since internal structures concerning structural validity (including unidimensionality), internal consistency, cross-cultural validity, and other forms of measurement invariance, which refer to how the different items in a PROM are related, are based on a reflective model and not a formative model, it is essential to identify the model used to develop the PROM (Mokkink et al., 2018). However, the studies did not report the kind of measurement model. According to the COSMIN user manual, if the measurement model is not reported clearly, an assessment of the internal structures is still recommended. As the studies (Garba et al., 2020; Gonzales, 2012; Minogue et al., 2018; Ourry et al., 2021; Paplikar et al., 2020; Relander et al., 2021; Valenzuela & Sachdev, 2007) did not report whether the instrument was based on a reflective, a formative or mixed measurement model, thus, we assessed all the related measurement properties in this review, but it is still critical to be cautious about the internal structure evaluated in this systematic review.

Several reliability tests, including test-retest, inter-rater, and intra-rater, were used to assess the reliability of the PROMs. All the PROMs evaluated in this review provided sufficient reliability indicating that the extent of cognitive reserve scores for the elderly was the same for repeated measurement under several conditions over time (De Vet et al., 2011). However, the evidence for reliability ranged from moderate (Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007) to low (Garba et al., 2020) or very low (Gonzales, 2012; Minogue et al., 2018) due to the high risk of bias and relatively low sample size. Therefore, the results of the reliability of cognitive reserve instruments remain possibly inconclusive. Studies with a larger sample size are recommended to improve the instrument's reliability. Furthermore, to lower the risk bias of the instrument's reliability in terms of study design, future studies should report the appropriateness of the time interval on the test-retest. As an aspect of statistical methods, ICC is recommended compared with Pearson or Spearman correlation coefficient for continuous scores, while weighted Kappa with a weighting scheme is suitable for ordinal scores, and Kappa for dichotomous and nominal scores.

Criterion validity can be subdivided into concurrent validity and predictive validity (De Vet et al., 2011). Although some studies stated that the criterion validity (Garba et al., 2020) or concurrent validity (Gonzales, 2012) of the instrument was assessed, hypotheses testing for convergent validity were performed. Firstly, convergent validity is based on the hypothesis principle whereby different

measures of the same construct should correlate highly with each other (De Vet et al., 2011). Secondly, concurrent validity considers both the score for the measurement and the gold standard simultaneously (De Vet et al., 2011). Nevertheless, the COSMIN panel concluded that no gold standard exists for PROMs (Mokkink et al., 2010), the only exception is when a shortened instrument is compared to the original long version. Based on this, the criterion validity of A-LEQ (Gonzales, 2012) should be compared with LEQ (Valenzuela & Sachdev, 2007). WTAR (Garba et al., 2020) and MMSE (Gonzales, 2012) were shown as gold standard measurement tools of cognitive reserve in these two studies, but IQ and cognitive function were not the reliable gold standard of cognitive reserve, because cognitive reserve has rarely been assessed directly, and functional brain process can be regarded as the closest direct measure (Stern et al., 2020). Thirdly, predictive validity is used to assess whether the measurement instrument predicts the gold standard in the future (De Vet et al., 2011), thus, predictive validity for cognitive reserve may predict less decline than expected given the degree of brain pathology present in the longitudinal studies. Given different operational definitions of cognitive reserve, various standards can be chosen as the closest to the gold standard but cannot be considered the gold standard. Therefore, the operational definitions of the cognitive reserve are very critical, which determines the psychometric properties.

Tool Recommendations

Evidence-backed summary of measurement properties can ultimately be used in providing recommendations for selecting the most appropriate instruments for cognitive reserve among older adults. To achieve this, the evidence for selecting suitable instruments in the included studies relays on the evaluation of the measurement properties, interpretability, and feasibility. Overall, we did not find any PROMs with high-quality evidence for an insufficient measurement property, and there were no studies with sufficient content validity at any quality level and sufficient internal consistency with evidence that is at least low-quality. Based on COSMIN recommendations for the most appropriate PROM, PROMs in this review are not rated as 'A' or 'C', therefore, they could be classified as 'B', which means being apt to be recommended for use, but demanding further research to assess the quality of these instruments. Furthermore, since there was not the best evidence for content validity, we cannot find the one with to be recommended provisionally.

In addition, selecting a suitable PROM depends not only on assessing the measurement properties but also on interpretability and feasibility. The LEQ (Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007) and RICE (Minogue et al., 2018) were specifically developed

for older adults and have been evaluated extensively. However, the completion time of LEQ is at least 30 min for older adults (Ourry et al., 2021; Paplikar et al., 2020; Valenzuela & Sachdev, 2007), which takes longer than E-CRIq. As a result, some older adults may need a break halfway to complete the questionnaire, which may increase the cognitive burden and lead to a long data collection duration (De Vet et al., 2011). In addition, the computation of the LEQ score is complex, as three sections of life stages were calculated at different points respectively, and a cumulative score for a specific section was summed and then normalized with the non-specific subscores, finally, each life stage was weighted 33.3% towards the overall LEQ total score. In contrast, the CRIq score can be easily calculated using an excel tool, which computes more conveniently.

Implications for Clinical Practice

From a clinical perspective, high cognitive reserve is a key factor in achieving cognitively healthy aging (Foley et al., 2012). The higher cognitive reserve can reduce the risk of mild cognitive impairment (MCI) and AD (Aguirre-Acevedo et al., 2016; Soldan et al., 2017; Xu et al., 2019), and affect the process of AD transformation (Bessi et al., 2018; Han et al., 2021; Xu et al., 2020). Based on this, an important consideration is the assessment of cognitive reserve using instruments among older adults, which is crucial in facilitating screening for older adults, especially those who are susceptible to or resistant to cognitive impairment due to aging or brain injury.

This systematic review suggests that all the included cognitive reserve instruments have the potential for assessment of cognitive reserve for older adults, it is helpful to provide an instrument basis for explaining the existing intervention mechanism and efficacy or effectiveness. Nowadays, various measurements may lead to some confusion about the choice of appropriate methods. Except for questionnaires, modeling cognitive reserve as a residual between risk factors, cognition, and brain status is a common method to quantify cognitive reserve as well (Bocancea et al., 2021), whereas the validity of the residual approach is determined by the variables, and easy to capture unexplained variance. Hence, valid and reliable questionnaires could facilitate the clinical assessment of cognitive reserve at a low cost and convenience to administer, which are standardized for clinical measurement to ensure comparability and generalizability between different groups and studies. Furthermore, these questionnaires method could be used as determinants of cognitive reserve and enable the computation of residual methods more objectively.

To hone in on the specific quality of these questionnaires, we used the up-to-date COSMIN to assess the quality of included cognitive reserve questionnaires among

older adults, and then summarize the quality of measurement properties, interpretability, and feasibility. This study supports evidence-based recommendations in selecting the most appropriate PROM for clinical practice, and different PROMs may be suitable for various purposes and may depend on feasibility aspects as well. Therefore, clinicians and nurses could choose optimal instruments to assess cognitive reserve among older adults considering the quality, time, and human conditions in a different context. If the time is not limited, LEQ will be a suitable instrument, conversely, E-CRIq is appropriate. In addition, the RICE was developed specifically for the Aboriginal Australian population, potentially more applicable to this population.

Implications for Research

Efforts to explore cognitive reserve have been hindered by measurement difficulties, some research gaps are worthy of further exploration. This systematic review of PROMs can identify certain pitfalls and gaps in previous studies about the measurement properties of the PROMs and offer more guidance about how future efforts to develop, validate, or adapt for cognitive reserve should be approached.

Future studies, digging into the validation of the questionnaires on older adults deeply, are needed. The exclusion of nine studies (Altieri et al., 2018; Amoretti et al., 2019; Apolinario et al., 2013; Çebi & Kulce, 2021; Kaur et al., 2021; León et al., 2014; Maiovis et al., 2016; Nucci et al., 2012; Ozakbas et al., 2021) that some of them were not designed for older adults, or did not measure cognitive reserve in the population of older adults, or did not validate questionnaires in older adults. Furthermore, the studies included in this review mainly focused on older adults without cognitive impairment. Only Garba et al. (2020) collected data from AD adults, and Ourry et al. (2021) included adults suffering from subjective cognitive decline. Since the significance of cognitive reserve for older adults, and the higher cognitive reserve was associated with a faster cognitive decline for older adults with MCI or AD (Soldan et al., 2017), thus, it would be valuable to design, adapt or validate instruments to assess cognitive reserve among the older population, including those with cognitive impairment.

Furthermore, the operational definition of the cognitive reserve was various, for example, there existed studies that showed the most complete paradigm of the cognitive reserve should include life experience information, cognitive tests, and magnetic resonance imaging analysis (Nogueira et al., 2022), thus, the construct validity of cognitive reserve hinges upon its ability to moderate the influence of brain pathology on cognition, and any assessment of a CR questionnaire's construct validity must involve testing the brain pathology and CR interaction effect on cognitive outcomes. While Stern et al. (2020) stated functional brain process, which was the operational definition in this review, is closer

to direct measures, such as functional magnetic resonance imaging. Cognitive reserve as the measurement characteristics must be defined clearly to reduce the risk of affecting the quality of the study design, which is a central issue to the measurement properties of cognitive reserve and should be made more explicit in future studies.

Content validity is the most vital measurement property that all other properties are hinged. However, content validity was lacking in the included studies. Therefore, for future studies, we recommend using the COSMIN to assess the content validity of the cognitive reserve instrument (Mokkink et al., 2019). For example, this aspect of development, validation, and adaptation of instruments needs to be improved by asking professionals about the relevance and comprehensiveness of a cognitive interview or focus group approach.

Formative and reflective measurements are easily ignored, which causes a persistent problem in neuropsychological measurement research. The items used to measure cognitive reserve were typically based on proxy indicators that contributed to cognitive reserve (Garba et al., 2020; Gonzales, 2012; Minogue et al., 2018; Ourry et al., 2021; Paplikar et al., 2020; Relander et al., 2021; Valenzuela & Sachdev, 2007). Therefore, cognitive reserve instruments were more likely to be based on a formative model or formative second-ordered mixed model. Hence, less correlation between the measurement items can be explained and understood. Furthermore, from the basis of the establishment of the scale, it is formative. But discuss it in depth, cognitive reserve and the proxies may exert to reverse causation (Stern et al., 2020). Therefore, we recommend that the development and testing of reserve theories regarding cognitive reserve require further research to clarify the measurement model choices (Jones et al., 2011). In addition, researchers should pay more attention to the difference between the reflective and formative models to avoid assessing unnecessary measurement properties and missing important measurement properties.

At the same time, more well-designed and rigorously conducted studies are required to assess cognitive reserve. On one hand, researchers might consider investigating instruments of cognitive reserve longitudinally, as the cognitive reserve is one of the predictors of cognitive trajectory in the elderly (Sardella et al., 2021; Vallet et al., 2020). Based on this, future research should continue to further validate the predictive validity in the elderly population, to further determine the clinical predictive value. Furthermore, longitudinal or experimental designs with larger sample sizes are warranted to undertake help validate cognitive reserve instruments guided by COSMIN methodology for PROMs (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018) and COSMIN Study Design checklist for PROM instruments (Mokkink et al., 2019).

Limitations

This review was subject to several limitations that should be noted. First, only English and Chinese studies were included. Therefore, potentially good instruments written in other languages were not included. For example, the CRQ (Rami et al., 2011) was developed in 2011 to assess cognitive reserve in healthy adults and validated in people with AD. This questionnaire can be completed within two minutes. However, it is only available in Spanish and Portuguese versions now. Although excluding these studies relatively limited our sample size, we chose to focus solely on cognitive reserve among older adults. Second, conference proceedings were also excluded, and therefore relevant new findings might be missing. Finally, it was not possible to pool the results in a quantitative meta-analysis due to the methodological heterogeneity of the limited available studies.

Conclusions

Validated instruments for assessing cognitive reserve in older adults are essential to facilitate screening and prevent or slow down the progression of cognitive decline. This systematic review has identified five instruments used to assess cognitive reserve in older adults, whereas the measurement error and criterion validity of the instruments were not evaluated in any of the studies. Additionally, most of the information regarding the measurement properties was incomplete. In summary, based on the current evidence for all measurement properties, all PROMs identified in this review have the potential to be used to measure cognitive reserve for older adults but lack a thorough validation of the content validity or appropriate interpretability and feasibility. Therefore, further studies are recommended to validate the measurement properties of existing cognitive reserve instruments for older adults according to the COSMIN methodology for PROMs (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018) and COSMIN Study Design checklist for PROM instruments (Mokkink et al., 2019).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11065-023-09594-3>.

Authors' Contributions Wanrui Wei: Conceptualization, Methodology, Validation, Formal analysis, Data Curation, Writing - Original Draft, Review & Editing, Project administration. Kairong Wang: Methodology, Validation, Formal analysis, Data Curation, Writing - Review & Editing. Jiyuan Shi: Methodology, Validation, Formal analysis, Data Curation, Writing - Review & Editing. Zheng Li: Supervision, Conceptualization, Methodology, Validation, Formal analysis, Data Curation, Writing - Review & Editing, Funding acquisition, Project administration.

Funding This work was supported by the School of Nursing of Peking Union Medical College [grant number: PUMCSON202101].

Data Availability All data generated or analyzed during this study are included in this published article and its supplementary files.

Declarations

Ethical Approval Not applicable.

Competing Interests The authors declare no potential conflicts of interest.

References

- Aguirre-Acevedo, D. C., Lopera, F., Henao, E., Tirado, V., Muñoz, C., Giraldo, M., et al. (2016). Cognitive decline in a Colombian kindred with autosomal dominant Alzheimer disease. *JAMA Neurology*, 73(4), 431. <https://doi.org/10.1001/jamaneurol.2015.4851>
- Almeida-Meza, P., Steptoe, A., & Cadar, D. (2021). Markers of cognitive reserve and dementia incidence in the English Longitudinal Study of Ageing. *British Journal of Psychiatry*, 218(5), 243–251. <https://doi.org/10.1192/bjp.2020.54>
- Altieri, M., Siciliano, M., Pappacena, S., Roldán-Tapia, M. D., Trojano, L., & Santangelo, G. (2018). Psychometric properties of the Italian version of the Cognitive Reserve Scale (I-CRS). *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 39(8), 1383–1390. <https://doi.org/10.1007/s10072-018-3432-0>
- Amoretti, S., Cabrera, B., Torrent, C., Bonnín, C. M., Mezquida, G., Garriga, M., et al. (2019). Cognitive Reserve Assessment Scale in Health (CRASH): Its validity and reliability. *Journal of Clinical Medicine*, 8(5), 586. <https://doi.org/10.3390/jcm8050586>
- Apolinario, D., Brucki, S. M. D., Ferretti, R. E., de Farfel, L., Magaldi, J. M., Busse, R. M., & Jacob-Filho, W. (2013). Estimating pre-morbid cognitive abilities in low-educated populations. *PLoS One*, 8(3), e60084. <https://doi.org/10.1371/journal.pone.0060084>
- Arenaza-Urquijo, E. M., & Vemuri, P. (2018). Resistance vs resilience to Alzheimer disease: Clarifying terminology for preclinical studies. *Neurology*, 90(15), 695–703. <https://doi.org/10.1212/WNL.0000000000005303>
- Arenaza-Urquijo, E. M., & Vemuri, P. (2020, April 14). Improving the resistance and resilience framework for aging and dementia studies. *Alzheimer's Research and Therapy*. BioMed Central Ltd. <https://doi.org/10.1186/s13195-020-00609-2>
- Bennett, D. A., Schneider, J. A., Arvanitakis, Z., Kelly, J. F., Aggarwal, N. T., Shah, R. C., & Wilson, R. S. (2006). Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology*, 66(12), 1837–1844. <https://doi.org/10.1212/01.wnl.0000219668.47116.e6>
- Bessi, V., Mazzeo, S., Padiglioni, S., Piccini, C., Nacmias, B., Sorbi, S., & Bracco, L. (2018). From subjective cognitive decline to Alzheimer's disease: The predictive role of neuropsychological assessment, personality traits, and cognitive reserve. A 7-year follow-up study. *Journal of Alzheimer's Disease*, 63(4), 1523–1535. <https://doi.org/10.3233/JAD-171180>
- Bocancea, D. I., van Loenhoud, A. C., Groot, C., Barkhof, F., van der Flier, W. M., & Ossenkoppele, R. (2021). Measuring resilience and resistance in aging and Alzheimer Disease using residual methods: A systematic review and Meta-analysis. *Neurology*, 97(10), 474–488. <https://doi.org/10.1212/WNL.00000000000012499>
- Boyle, R., Knight, S. P., De Looze, C., Carey, D., Scarlett, S., Stern, Y., et al. (2021). Verbal intelligence is a more robust cross-sectional measure of cognitive reserve than level of education in healthy older adults. *Alzheimer's Research & Therapy*, 13(1), 128. <https://doi.org/10.1186/s13195-021-00870-z>
- Çebi, M., & Kulce, S. N. (2021). The Turkish translation study of the Cognitive Reserve Index Questionnaire (CRIQ). *Applied Neuropsychology: Adult*. <https://doi.org/10.1080/23279095.2021.1896519>
- Choi, C. H., Park, S., Park, H. J., Cho, Y., Sohn, B. K., & Lee, J. Y. (2016). Study on Cognitive Reserve in Korea using Korean Version of Cognitive Reserve Index Questionnaire. *Journal of Korean Neuropsychiatric Association*, 55(3), 256. <https://doi.org/10.4306/jknpa.2016.55.3.256>
- Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, 61(12), 1250–1262. <https://doi.org/10.1016/j.jbusres.2008.01.013>
- De Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge University Press.
- Foley, J. M., Ettenhofer, M. L., Kim, M. S., Behdin, N., Castellon, S. A., & Hinkin, C. H. (2012). Cognitive reserve as a protective factor in older HIV-positive patients at risk for cognitive decline. *Applied Neuropsychology Adult*, 19(1), 16–25. <https://doi.org/10.1080/09084282.2011.595601>
- Garba, A. E., Grossberg, G. T., Enard, K. R., Jano, F. J., Roberts, E. N., Marx, C. A., & Buchanan, P. M. (2020). Testing the Cognitive Reserve Index questionnaire in an Alzheimer's disease population. *Journal of Alzheimer's Disease Reports*, 4(1), 513–524. <https://doi.org/10.3233/adr-200244>
- Gonzales, C. (2012). *The psychometric properties of the lifetime experience questionnaire (LEQ) in older American adults*. ProQuest Dissertations and Theses. The University of North Carolina at Greensboro. Retrieved January 10, 2022, from <https://www.proquest.com/dissertations-theses/psychometric-properties-lifetime-experience/docview/1022181255/se-2?accountid=205795>
- Griffiths, C., Armstrong-James, L., White, P., Rumsey, N., Pleat, J., & Harcourt, D. (2015). A systematic review of patient reported outcome measures (PROMs) used in child and adolescent burn research. *Burns: Journal of the International Society for Burn Injuries*, 41(2), 212–224. <https://doi.org/10.1016/j.burns.2014.07.018>
- Han, S., Hu, Y., Pei, Y., Zhu, Z., Qi, X., & Wu, B. (2021). Sleep satisfaction and cognitive complaints in chinese middle-aged and older persons living with HIV: The mediating role of anxiety and fatigue. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*, 33(7), 929–937. <https://doi.org/10.1080/09540121.2020.1844861>
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2), 199–218. <https://doi.org/10.1086/376806>
- Jones, R. N., Manly, J., Glymour, M. M., Rentz, D. M., Jefferson, A. L., & Stern, Y. (2011). Conceptual and Measurement Challenges in Research on Cognitive Reserve. *Journal of the International Neuropsychological Society*, 17(4), 593–601. <https://doi.org/10.1017/S1355617710001748>
- Kartschmit, N., Mikolajczyk, R., Schubert, T., & Lacruz, M. E. (2019). Measuring cognitive reserve (CR) – a systematic review of measurement properties of CR questionnaires for the adult population. *PLoS One*, 14(8), 1–23. <https://doi.org/10.1371/journal.pone.0219851>
- Kaur, N., Fellows, L. K., Brouillette, M. J., & Mayo, N. (2021). Development and validation of a cognitive reserve index in HIV. *Journal of the International Neuropsychological Society*, (2021), 1–9. <https://doi.org/10.1017/S1355617721000461>
- Landenberger, T., Cardoso, N. O., de Oliveira, C. R., & de Argimon, I. I. L. (2019). Instruments for measuring cognitive reserve: A systematic review. *Psicologia - Teoria e Prática*, 21(2). <https://doi.org/10.5935/1980-6906/psicologia.v21n2p58-74>

- Lee, J., Lee, E. H., Chae, D., & Kim, C. J. (2020). Patient-reported outcome measures for diabetes self-care: A systematic review of measurement properties. *International Journal of Nursing Studies*, *105*, 103498. <https://doi.org/10.1016/j.ijnurstu.2019.103498>
- León, I., García-García, J., & Roldán-Tapia, L. (2014). Estimating cognitive reserve in healthy adults using the cognitive reserve scale. *PLoS One*, *9*(7), e102632. <https://doi.org/10.1371/journal.pone.0102632>
- León, I., García, J., & Roldán-Tapia, L. (2011). Development of the scale of cognitive reserve in spanish population: A pilot study. *Revista de Neurologia*, *52*(11), 653–660. <https://doi.org/10.33588/rn.5211.2010704>
- Leon-Estrada, I., Garcia-Garcia, J., & Roldan-Tapia, L. (2017). [Cognitive Reserve Scale: Testing the theoretical model and norms]. *Revista De Neurologia*, *64*(1), 7–16.
- Li, X., Song, R., Qi, X., Xu, H., Yang, W., Kivipelto, M., et al. (2021). Influence of cognitive reserve on cognitive trajectories: Role of brain pathologies. *Neurology*, *97*(17), e1695–e1706. <https://doi.org/10.1212/WNL.00000000000012728>
- Maiovis, P., Ioannidis, P., Nucci, M., Gotzamani-Psarrakou, A., & Karacostas, D. (2016). Adaptation of the Cognitive Reserve Index Questionnaire (CRIq) for the greek population. *Neurological Sciences*, *37*(4), 633–636. <https://doi.org/10.1007/s10072-015-2457-x>
- Minogue, C., Delbaere, K., Radford, K., Broe, T., Forder, W. S., & Lah, S. (2018). Development and initial validation of the Retrospective Indigenous Childhood Enrichment scale (RICE). *International Psychogeriatrics*, *30*(4), 519–526. <https://doi.org/10.1017/S104161021700179X>
- Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, *10*(1), 22. <https://doi.org/10.1186/1471-2288-10-22>
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2019). COSMIN study design checklist for patient-reported outcome measurement instruments. www.cosmin.nl. Accessed 17 Apr 2022
- Morris, J. C., Storandt, M., McKeel, D. W., Rubin, E. H., Price, J. L., Grant, E. A., & Berg, L. (1996). Cerebral amyloid deposition and diffuse plaques in “normal” aging: Evidence for presymptomatic and very mild Alzheimer’s disease. *Neurology*, *46*(3), 707–719. <https://doi.org/10.1212/WNL.46.3.707>
- Negash, S., Wilson, R., Leurgans, S., Wolk, D., Schneider, J., Buchman, A., et al. (2013). Resilient brain aging: Characterization of discordance between Alzheimer’s disease pathology and cognition. *Current Alzheimer Research*, *10*(8), 844–851. <https://doi.org/10.2174/15672050113109990157>
- Nelson, M. E., Jester, D. J., Petkus, A. J., & Andel, R. (2021). Cognitive reserve, Alzheimer’s neuropathology, and risk of dementia: A systematic review and meta-analysis. *Neuropsychology Review*, *31*(2), 233–250. <https://doi.org/10.1007/s11065-021-09478-4>
- Nogueira, J., Gerardo, B., Santana, I., Simões, M. R., & Freitas, S. (2022). The assessment of cognitive reserve: a systematic review of the most used quantitative measurement methods of cognitive reserve for aging. *Frontiers in Psychology*, *13*(March), 1–9. <https://doi.org/10.3389/fpsyg.2022.847186>
- Nucci, M., Mapelli, D., & Mondini, S. (2012). Cognitive Reserve Index questionnaire (CRIq): A new instrument for measuring cognitive reserve. *Aging Clinical and Experimental Research*, *24*(3), 218–226. <https://doi.org/10.3275/7800>
- Opdebeeck, C., Martyr, A., & Clare, L. (2016). Cognitive reserve and cognitive function in healthy older people: A meta-analysis. *Neuropsychology Development and Cognition Section B Aging Neuropsychology and Cognition*, *23*(1), 40–60. <https://doi.org/10.1080/13825585.2015.1041450>
- Ourry, V., Marchant, N. L., Schild, A. K., Coll-Padros, N., Klimecki, O. M., Krolak-Salmon, P., et al. (2021). Harmonisation and between-country differences of the lifetime of Experiences Questionnaire in older adults. *Frontiers in Aging Neuroscience*, *13*(October), 1–12. <https://doi.org/10.3389/fnagi.2021.740005>
- Ozakbas, S., Yigit, P., Akyuz, Z., Sagici, O., Abasiyanik, Z., Ozdogar, A. T. (2021). Validity and reliability of “Cognitive Reserve Index Questionnaire” for the Turkish Population. *Multiple Sclerosis and Related Disorders*, *50*(June 2020). <https://doi.org/10.1016/j.msard.2021.102817>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. <https://doi.org/10.1136/bmj.n71>
- Paplikar, A., Ballal, D., Varghese, F., Sireesha, J., Dwivedi, R., Rajan, A., et al. (2020). Assessment of lifestyle experiences across lifespan and cognitive ageing in the Indian context. *Psychology and Developing Societies*, *32*(2), 308–330. <https://doi.org/10.1177/0971333620937512>
- Patterson, C. (2018). *World Alzheimer report 2018 - The state of the art of dementia research: New frontiers*. Alzheimer’s Disease International.
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Rami, L., Valls-Pedret, C., Bartrés-Faz, D., Caprile, C., Solé-Padullés, C., Castellví, M., et al. (2011). Cognitive reserve questionnaire. Scores obtained in a healthy elderly population and in one with Alzheimer’s disease. *Revista de Neurologia*, *52*(4), 195–201. <https://doi.org/10.33588/rn.5204.2010478>
- Relander, K., Mäki, K., Soenne, L., García-García, J., & Hietanen, M. (2021). Active lifestyle as a reflection of cognitive reserve: The modified Cognitive Reserve Scale. *Nordic Psychology*, *73*(3), 242–252. <https://doi.org/10.1080/19012276.2021.1902846>
- Sajeev, G., Weuve, J., Jackson, J. W., VanderWeele, T. J., Bennett, D. A., Grodstein, F., & Blacker, D. (2016). Late-life cognitive activity and dementia: A systematic review and bias analysis. *Epidemiology (Cambridge, Mass.)*, *27*(5), 732–742. <https://doi.org/10.1097/EDE.0000000000000513>
- Sardella, A., Quattropiani, M. C., & Basile, G. (2021). Can cognitive reserve protect frail individuals from dementia? *The Lancet Healthy Longevity*, *2*(2), e67. [https://doi.org/10.1016/S2666-7568\(20\)30070-2](https://doi.org/10.1016/S2666-7568(20)30070-2)
- Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (2013). GRADE Handbook - Introduction to GRADE Handbook: Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. *The GRADE Working Group*. <https://gdt.gradeapro.org/app/handbook/handbook.html>. Accessed 4 May 2022
- Soldan, A., Pettigrew, C., Cai, Q., Wang, J., Wang, M., Moghekar, A., et al. (2017). Cognitive reserve and long-term change in cognition in aging and preclinical Alzheimer’s disease. *Neurobiology of Aging*, *60*, 164–172. <https://doi.org/10.1016/j.neurobiolaging.2017.09.002>

- Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., et al. (2020). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's and Dementia*. <https://doi.org/10.1016/j.jalz.2018.07.219>
- Sumowski, J. F., Rocca, M. A., Leavitt, V. M., Dackovic, J., Mesaros, S., Drulovic, J., et al. (2014). Brain reserve and cognitive reserve protect against cognitive decline over 4.5 years in MS. *Neurology*, *82*(20), 1776–1783. <https://doi.org/10.1212/WNL.0000000000000433>
- Terwee, C. B., Jansma, E. P., Riphagen, I. I., & De Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, *18*(8), 1115–1123. <https://doi.org/10.1007/s11136-009-9528-5>
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., et al. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, *27*(5), 1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>
- Valenzuela, M. J., & Sachdev, P. (2007). Assessment of complex mental activity across the lifespan: Development of the lifetime of Experiences Questionnaire (LEQ). *Psychological Medicine*, *37*(7), 1015–1025. <https://doi.org/10.1017/S003329170600938X>
- Valenzuela, M., Brayne, C., Sachdev, P., Wilcock, G., & Matthews, F. (2011). Cognitive lifestyle and long-term risk of dementia and survival after diagnosis in a multicenter population-based cohort. *American Journal of Epidemiology*, *173*(9), 1004–1012. <https://doi.org/10.1093/aje/kwq476>
- Vallet, F., Mella, N., Ihle, A., Beaudoin, M., Fagot, D., Ballhausen, N., et al. (2020). Motivation as a mediator of the relation between cognitive reserve and cognitive performance. *The journals of gerontology Series B*, *75*(6), 1199–1205. <https://doi.org/10.1093/geronb/gby144>
- Van Loenhoud, A. C., Van Der Flier, W. M., Wink, A. M., Dicks, E., Groot, C., Twisk, J., et al. (2019). Cognitive reserve and clinical progression in Alzheimer disease: A paradoxical relationship. *Neurology*, *93*(4), E334–E346. <https://doi.org/10.1212/WNL.00000000000007821>
- Wilson, R. S., Barnes, L. L., & Bennett, D. A. (2003). Assessment of lifetime participation in cognitively stimulating activities. *Journal of Clinical and Experimental Neuropsychology*, *25*(5), 634–642. <https://doi.org/10.1076/jcen.25.5.634.14572>
- Xu, H., Yang, R., Qi, X., Dintica, C., Song, R., Bennett, D. A., & Xu, W. (2019). Association of lifespan cognitive reserve indicator with dementia risk in the presence of brain pathologies. *JAMA Neurology*, *76*(10), 1184–1191. <https://doi.org/10.1001/jamaneurol.2019.2455>
- Xu, H., Yang, R., Dintica, C., Qi, X., Song, R., Bennett, D. A., & Xu, W. (2020). Association of lifespan cognitive reserve indicator with the risk of mild cognitive impairment and its progression to dementia. *Alzheimer's and Dementia*, *16*(6), 873–882. <https://doi.org/10.1002/alz.12085>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.