Psychometric Properties of Cognitive Screening Tools in Brain Injury

Jessica McLaren (BSc, MSc)

Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology

Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University of Glasgow

September 2022

Table of Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to give a huge thanks to my supervisor Dr Breda Cullen, for being with me every step of the way during my research. Your expertise, support and quick responses to all of my many, many research queries; have been absolutely invaluable. I am grateful to have had such a fantastic supervisor. Thank you also to Dr Hamish McLeod for casting a critical eye over my research and helping me along through the process.

Thank you to Dr Brain O'Neill and Dr Sara Da Silva Ramos at the Brain Injury Rehabilitation Trust. To Dr O'Neill for helping to conceptualise and support the initial development of the project; and to Dr Da Silva Ramos for kindly offering to support the project to completion. Thank you to the staff at Graham Anderson House, who helped collate data under challenging circumstances during the pandemic. A special thank you also to the patients at Graham Anderson House, without whom this project would not have been possible.

Thank you to Paul Cannon, your library expertise and calm demeanour helped me make sense of the perplexing world of systematic review search strategies. Thank you also to my colleague, Alex Fradera, for being such a patient, thorough and considerate second marker. You (and your cat joining our zoom calls) made the process so much more enjoyable.

Thank you to my Mum, Dad and sisters; you have supported me every step of the way over the years. To my fiancé Marc; thank you for comforting me at every set-back and stressful moment; and for making sure I was well supplied with coffee and cake. To my friends, I am forever grateful to you all for all the support you gave me.

Finally, to the class of 2022. Not only did we make it through the Doctorate, we did it during the first three years of the COVID-19 pandemic. Despite the distance between us; through zoom calls, phonecalls and texts, we comforted each other and reassured each other that we weren't going through the process alone. I will forever be grateful to our resilient, brilliant cohort.

Foreword

Data collection for this project was conducted by staff at Graham Anderson House, and began in 2013. It was reported in April 2022 that the estimated number of recorded cognitive screening data was lower than initially anticipated in the service database. Despite best efforts by staff, short-staffing and COVID-prevention protocols then slowed additional data collation and extraction within the service. As a result, Chapter 2 of this thesis is being submitted as a COVID-19 contingency project (Option 1).

Chapter 1: Systematic Review

The Reliability and Validity of Brief Cognitive Screening Tools used in Traumatic Brain Injury: a Systematic Review

Authors:

Jessica McLaren[1*], Alex Fradera[2], Dr Breda Cullen[3]

1. Mental Health and Wellbeing, Institute of Health and Wellbeing, University of Glasgow, Scotland, United Kingdom; and NHS Ayrshire and Arran, Scotland, United Kingdom
2. Mental Health and Wellbeing, Institute of Health and Wellbeing, University of Glasgow, Scotland, United Kingdom
3. Mental Health and Wellbeing, Institute of Health and Wellbeing, University of Glasgow, Scotland, United Kingdom;and NHS Lanarkshire, Scotland, United Kingdom

Corresponding author: j.mclaren.2@research.gla.ac.uk

Prepared in accordance with Frontiers in Psychology: Neuropsychology

https://www.frontiersin.org/journals/psychology/for-authors/author-guidelines

Word Count: 10943

Abstract

Reliable and valid cognitive screening tools are essential in Traumatic Brain Injury (TBI). Yet, there are no existing systematic reviews on the subject, nor is there a consensus about which tool should be used in clinical practice. This systematic review assessed psychometric properties of cognitive screening tools for detecting cognitive impairment in TBI. Inclusion criteria were: full-length articles published in peer-reviewed journals; with a sample of adults aged 18-80 diagnosed with TBI (any stage or severity); whose primary focus was validating a cognitive screening tool; with psychometrics consistent with COSMIN guidelines. Exclusion criteria included; studies with clinical populations other than TBI in the same analysis with people with TBI. Published literature was retrieved from: MEDLINE, Web of Science Core Collection, EMBASE, CINAHL and PsycINFO until the date of extraction (27/1/22). A narrative synthesis was performed. 33 studies evaluated the psychometric properties of a total of 18 cognitive screening tools, in a variety of languages. Types of validity assessed included structural validity, internal consistency, reliability, criterion validity (or diagnostic test accuracy), convergent/divergent validity and discriminant validity. The Montreal Cognitive Assessment (MoCA) and the Mini Mental State Exam (MMSE) were the most widely validated cognitive screening tools for use in TBI. The MoCA had the most promising evidence of its psychometric properties, which has implications for clinical practice. Future research should aim to follow standard criteria for psychometric studies to allow meaningful comparisons across the literature.

Word Count: 234

Keywords: Traumatic Brain Injury; Screening Tool; Cognitive Assessment; Neuropsychological Test; Reliability; Validity; Psychometric Assessment

Introduction

Traumatic brain injuries (TBI) are a prevalent cause of hospital admission, with 155,919 cases of head injury presenting to UK hospitals annually (Headway, 2017). Cognitive impairments in TBI are common in those with moderate-severe TBI, and in acute phase of mild TBI (Barman et al., 2016). A complete neuropsychological assessment is a high-quality method of identifying cognitive impairment but is resource intensive. Given the prevalence of TBI and the demands on services, there is a need for brief, valid and reliable screening tools. Screening tools have the potential to identify cognitive impairment in an efficient and cost-effective way, allowing the targeting of additional assessment, intervention, and, at a service level, could help plan resource allocation (NICE, 2014; Scottish Aquired Brain Injury Network, 2017; Teager et al., 2020).

A previous literature review was identified, but used a broader clinical population definition, including patients diagnosed with cerebrovascular accidents (Canadian Agency for Drugs and Technologies in Health (CADTH), 2014). Other systematic reviews have been published on cognitive screening tool use in stroke populations (Stolwyk et al., 2014; Kosgallana et al., 2019). However, important differences exist between stroke and TBI populations in pathology and the most prevalent profiles of cognitive impairment. Zhang *et al.*, (2016) compared the validity of two different cognitive screening tools in TBI and Stroke populations. The sensitivity of these tools to identify cognitive impairment differed between TBI and Stroke groups; suggesting the psychometric properties of cognitive screening tools should not be assumed to be consistent across brain injury groups. No systematic review or meta-analysis could be identified on the psychometric properties of cognitive screening tools, specifically in TBI populations.

A variety of cognitive screening tools exist which may be helpful for detecting cognitive impairment in TBI (Cullen et al., 2007; Abd Razak et al., 2019). However, the heterogeneity of cognitive impairments in TBI has presented a challenge to researchers to identify valid and reliable tools (Teager et al., 2020). A number of individual studies could be identified which reported on the validity of certain cognitive screening tools, such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Examination (MoCA), in TBI populations (Gaber, 2008; Zhang et al., 2016). These tools were initially developed for and are validated

for use in dementia populations. Major trauma centres often use widely available tools to assess TBI patients, such as the Addenbrookes Cognitive Exam-III (Hsieh et al., 2013) and the MoCA (Nasreddine et al., 2005; Teager et al., 2020), when the psychometric properties of these tools are either unknown or are yet to be summarised in a meaningful way. Therefore, there is a need to review the evidence for the psychometric properties of cognitive screening tools in TBI. This will support the refinement of clinical guidelines, which in turn, would contribute to evidence-based practice.

Objectives

This systematic review aimed to determine the validity and reliability of cognitive screening tools for detecting cognitive impairment in TBI populations. Research questions were:

1. How reliable are the screening tools used in TBI populations?
2. How valid are the screening tools used in TBI populations?

A secondary objective was to determine the comparative validity and reliability of identified screening tools between different severities (mild and moderate/severe) of TBI.

Methods

Protocol and registration

This systematic review was written in accordance with PRISMA (Page *et al.*, 2021; Appendix 1.1). The protocol was registered on PROSPERO on 11/01/22, and amended on 27/7/22 with additional exclusion criteria (Registration number CRD42022297346 [available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=297346].)

Eligibility Criteria

All eligible studies with an English language version, up to the search date (27/1/22) were included. For additional information inclusion and exclusion information see Appendix 1.2.

During screening and extraction, eligibility criteria were clarified to account for unanticipated nuances in the literature.

Patient Population

Studies were required to include human adults (aged 18-80) diagnosed with TBI of any severity or stage (mild, moderate or severe; acute or post-acute). The TBI may have been sustained when the individual was <18 years old. TBI was defined as an injury to the brain caused by an external force. Mixed samples which included conditions other than TBI (including other acquired brain injuries [ABI] such as stroke) in the same analysis with people with TBI, were excluded. If the study included a mixed sample of children and adults, a separate analysis with adults 18-80 only was required. Studies with adult participants which did not specify an upper or lower age limit were included.

Index Test

Studies were required to include a cognitive screening test or tool, using the definition outlined in Cullen et al., (2007); the test must have been designed to screen for cognitive impairment or be used for that purpose, have an administration time of less than 20 minutes and be available in English. Screens were required to be administered to patients and directly measure their cognitive performance. The screen could assess multiple domains or a single domain of cognitive function. Screening tests were excluded which were measures of, or were being used to measure: functional ability (including driving); malingering, effort or performance validity; and consciousness, lower level, subcortical or sensory functions (including basic visuo-ocular, vestibular, or auditory function). "Modified" or non-English language versions were noted in the report.

Comparator Test

Studies which included a comparator test (i.e. any other screening test that may have been used in the study for comparison) were included, but this was not a requirement.

Outcomes of Interest

A variety of psychometric outcomes were accepted depending on the study design. Accepted psychometric qualities and their measurement properties were consistent with those outlined in the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines (Mokkink et al., 2010). Examples of permitted psychometric qualities, designs and statistics include:

- Criterion Validity (or diagnostic accuracy); ability to detect cognitive impairment (measured by a standard neuropsychological test and appropriate reference criteria, as defined by the authors) or traumatic brain injury diagnosis. Relevant statistics may include: correlation, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the curve (AUC).
- Discriminant validity: ability to distinguish between TBI and control groups, or between severity of TBI status (mild, moderate, severe). Relevant statistics may include: independent t-tests with effect size statistic.
- Convergent or Divergent Validity: correlation with a measure expected to be related or divergent to the cognitive screening tool (such as a functional measure), where both measures are completed at the same time.
- Inter-rater reliability: relevant statistics may include intra-class correlations [ICCs].
- Test-retest reliability: ONLY where the researchers have deliberately administered the measure twice in close succession, within a baseline phase during which natural change is unlikely to have occurred (e.g. consecutive days within the first week in a rehab ward, or during the baseline phase of an intervention study before the intervention was introduced). Relevant statistics include: intra-class correlations [ICCs].
- Internal consistency: relevant statistics include Cronbach's alpha.

In addition to the above outcomes: information on measurement error, structural validity, cross-cultural validity and responsiveness (only where a standard cognitive test was used as a comparator) in accordance with COSMIN (Mokkink et al., 2010). Outcomes related to clinical utility, feasibility or acceptability were recorded where available. These outcomes were all of equal interest and none of which were prioritised.

Study Type

Included reports were full-length articles (in a peer reviewed journal) reporting primary research (not systematic reviews or conference abstracts etc.). Eligible studies assessed the psychometric properties of a cognitive screening tool, in accordance with the outcomes of interest, and had a group design (i.e. an observational or experimental design, but not a single case study or case series design). Examples of excluded designs were:

- the screening tool is compared to a self-report measure or brain injury measure or across different groups; but is being used as an outcome measure, there is no clear psychometric relevance or it is not being validated as a screen
- A cognitive screening tool is used to track improvement in an intervention or recovery study (where no additional tool is used to compare the screen to a gold-standard or any additional relevant psychometrics are calculated)
- The tool is evaluated in conditions not ecologically valid for clinical use
- psychometrics are present but are focused on an aspect of the tool other than cognitive function (e.g. performance validity)

Information Sources, Search Strategy and Study Selection

Scoping searches were conducted to refine search terms. Key papers identified in scoping were noted. The sensitivity of the search strategy was evaluated by its ability to detect the key papers. The full search strategy is outlined in Appendix 1.2. No additional limits were used in search filters. Published literature was retrieved from MEDLINE, Web of Science Core Collection, EMBASE, CINAHL, and PsycINFO, from inception up until the date of extraction (27/1/22). Search results were managed in the first author's RefWorks library (www.refworks.com/refworks2/). Duplicates were removed during database extraction. Information of interest was extracted into an Excel spreadsheet and decision making was recorded.

Titles and/or abstracts were screened by JM for eligibility. The second reviewer (AF) independently repeated this process for 100 records to check for consistency. For the second

screening phase, the full text was read by JM for all records identified as “maybe eligible” at the title/abstract stage and a decision was made about its eligibility. AF independently repeated this process for 20 records to check for consistency.

The reference lists of eligible papers and relevant systematic reviews were then searched by JM by hand (backward citation). Subsequent papers which have cited eligible papers, identified electronically using the "cited by" function in Google Scholar, were searched (forward citation) by JM on 10th July 2022. No additional studies or data were sought by contacting authors.

Data Collection and Data Items

Three data extraction templates were created to extract relevant data from eligible studies. Data extracted included: patient demographics (sample size, age, sex, TBI severity, time since injury; for both TBI and relevant control groups); cognitive screens (tool name, original reference, domains assessed, items, range of scores, time to administer etc.) and psychometric properties (including psychometric quality assessed, details of the study design and relevant measurement property; as defined in COSMIN). Additional information was gathered, including: inclusion/exclusion criteria, TBI diagnosis criteria and study setting, but was not reported for brevity. JM completed all data extraction and AF checked extraction for five of the papers. Only minor formatting and spelling errors were identified that would have been recognised by the primary author at the synthesising stage.

Risk of Bias

The COSMIN Risk of Bias checklist was used by JM to assess the quality of all included studies (Mokkink et al., 2010). It is a consensus-based checklist for evaluating the methodological quality of psychometric studies. Risk of bias was rated as: ‘very good’ (V), ‘adequate’ (A), ‘doubtful’ (D), ‘inadequate’ (I) or ‘not applicable’ (N) according to COSMIN criteria. A ‘worst counts score’ approach was used for overall rating. Additional details are given in Appendix 1.3. Risk of bias (for any relevant measurement properties) was assessed independently for five studies by AF.

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool was used to assess the quality, only for studies which specifically claimed to have a "diagnostic accuracy" design (or similar)(Whiting et al., 2011). It includes items which assess both the risk of bias and applicability of results. JM and AF independently rated the quality of three studies.

Summary Measures and Synthesis of Results

The evidence was graded using the COSMIN checklist for good measurement properties by JM. Measurement properties were rated as (+) sufficient; (-) insufficient; or (?) indeterminate, using criteria defined in COSMIN, which varied for each type of design (for example, criterion validity designs should report an area under the curve [AUC] ≥0.7 for a sufficient (+) rating (see Appendix 1.3). A narrative synthesis was performed, with data presented in text, tables and figures. Each study was summarised including a description of the demographics. Details of cognitive screening tool(s) used were reported. The type(s) of validity and reliability, relevant details of the study design and the resulting psychometric statistic(s) were reported. Measurement properties of the tools in TBI populations were discussed, considering the quality of the studies. While COSMIN recommends a Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach to synthesise the evidence across studies, given the expected heterogeneity of the literature identified in scoping searches, it was considered not feasible for this review.

Results

The initial title-abstract calibration between JM and AF resulted in "fair" agreement (81.18% agreement: Cohen's Kappa=0.29). After ambiguities within the criteria were clarified, 100% agreement was reached. Independent full-text screening resulted in 82.4% agreement between reviewers (Cohen's Kappa=0.62). After further discussion, this resolved to 88.88% agreement (Cohen's Kappa=0.73).

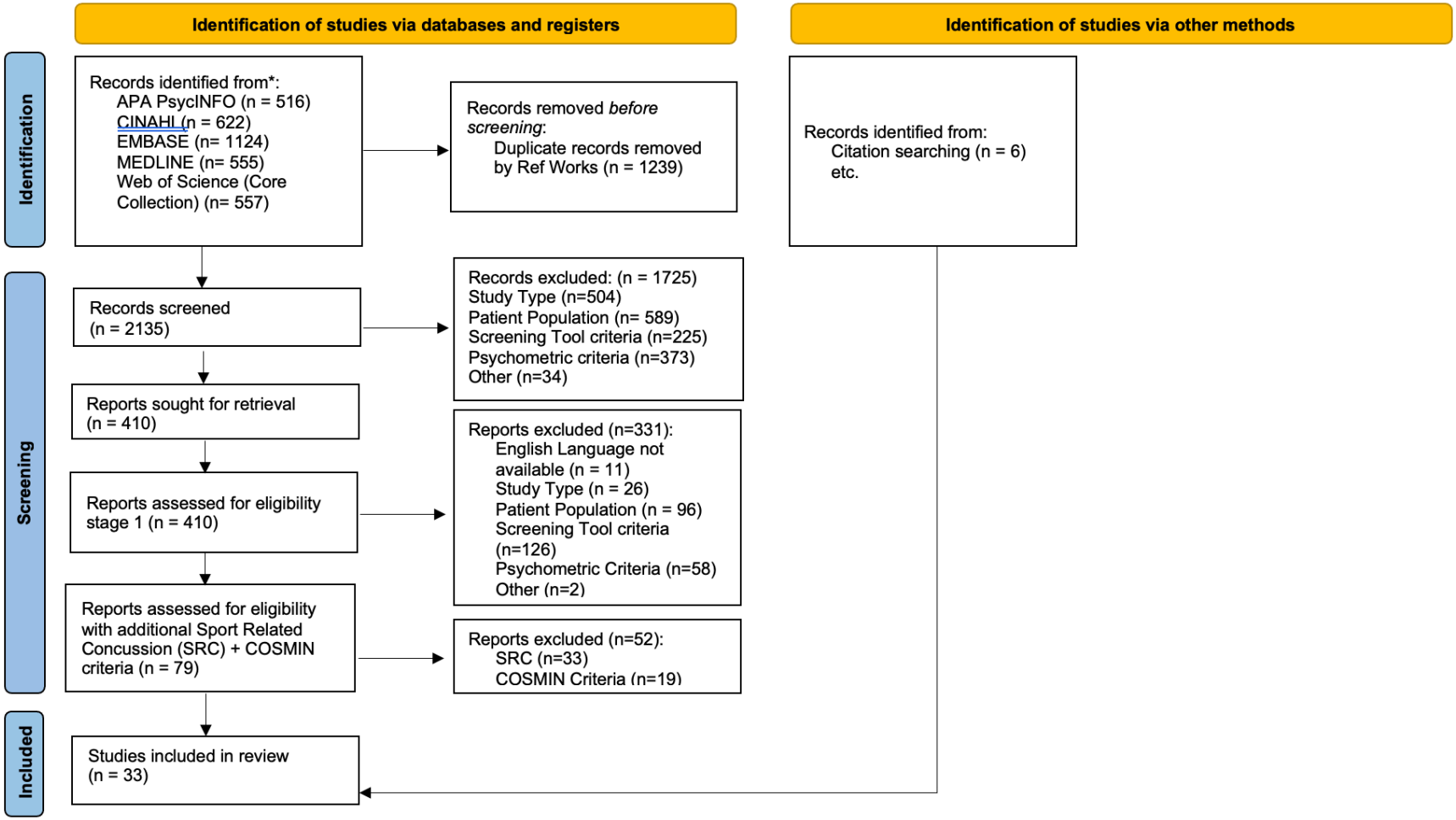
79 studies were initially deemed eligible for inclusion (see PRISMA Flow Chart in Figure1.1). Several studies which might have been expected to be included were found to be ineligible; including Gaber, (2008) and Hazan *et al.*, (2017), whose inclusion criteria did not meet the age criteria. After full-text screening, additional exclusion criteria were applied to account for unanticipated nuances in the literature:

1. Sports Related Concussion (SRC) Criteria: studies with a SRC sample were excluded. These studies often used measures designed for use in SRC settings (i.e. at pitch-side, in the acute phase, integrating neurological exam elements). Results were often compared to "baseline' screening, conducted pre-morbidly. Given these factors, it was felt this area of literature was distinct, less applicable to clinical settings; and would increase the heterogeneity of included studies.
2. COSMIN Criteria: COSMIN outlines that studies should be excluded if they "only use the PROM as an outcome measurement instrument…" or "… studies in which the PROM is used in a validation study of another instrument", as screening would be "extremely time consuming" without this criterion (Mokkink et al., 2010, pp 20). Therefore, studies were excluded where the purpose of the study was not to validate the screen and/or where the screen is used as an outcome measure or to validate another tool. Studies which did not have a design consistent with COSMIN (e.g. using multiple regression) were excluded.

After adding these criteria, and with the addition of six papers for citation searching; 33 studies were eligible for inclusion, in which 18 cognitive screening tools were evaluated. Sample characteristics and cognitive screens are outlined in Table 1.1 and Table 1.2 respectively. Demographic and screening tool metrics sought by this review were not consistently reported, and missing data indicates this information was not available. Information not reported was sought from referenced sources within the paper where available. The studies were heterogeneous in terms of TBI severity (12 mild or "concussion"; 6 moderate-severe; 11 mixed; and 4 unclear) and time since diagnosis (mean range: 5.6hrs to 60.33 months).

Of the 18 tools, administration time ranged between 2 minutes to 20 minutes (where reported). Some measures were designed as screening tools e.g. MoCA; others were designed as brief tests and were being evaluated as a cognitive screen e.g. COntrolled Word Association Test

(COWAT). Some evaluated multiple cognitive domains e.g. MMSE; others focused on one area of cognitive function e.g. Bethesda Eye and Attention Measure (BEAM).



Exclusion Criteria: No English Language Version Available (1); Study Type (2); Patient Population (3); Screening Tool Criteria (4); Psychometric criteria not met (5); Other (6) e.g. duplicate not previously identified by Refworks or a retracted study; Sports Related Concussion sample (7); COSMIN criteria not met (8).

Figure 1. 1 *PRISMA flowchart displaying excluded and included studies at each stage of the review process (Page et al., 2021)*

Risk of Bias

COSMIN Risk of Bias (RoB) ratings are reported within the psychometric results tables 1.3-1.5. Initial agreement in RoB ratings between independent reviewers was low (<50%). Key areas of ambiguity were clarified, and after independent rating for an additional three papers, 100% agreement was reached.

While RoB relating to structural validity, criterion validity and internal consistency was generally rated as "very good", issues were raised within reliability and divergent validity designs. Of note, many studies did not report an effect size or correlation statistic for their divergent validity analyses; giving an inadequate understanding of the magnitude of difference between groups. Other issues included: insufficient statistical analyses to capture the multidimensionality of measures, limited information of diagnosis procedures, the time delay between reliability measurements and poor reporting of statistical information.

Independent rating using the QUADAS 2 tool resulted in 76% agreement between the reviewers; and differences in opinion were discussed until consensus was reached. The results (Table 1.6) revealed that all diagnostic accuracy studies had a source of potential bias. In most studies, examiners were not blinded to the results of the patient status, participants were recruited into pre-determined groups based on diagnosis status and exploratory cut-offs were used in ROC curve analyses; all of which are potential sources of bias, as outlined in QUADAS.

Measurement Properties

Psychometric properties of the tools are reported in Tables 1.3-1.5. No studies measured cross cultural validity, measurement invariance or measurement error or responsiveness (in a way that was suggested in the COSMIN guidelines and which could be evaluated using their criteria). There was significant heterogeneity across studies in types of psychometric qualities assessed and subsequent outcome statistic. Divergent validity was the most common type of validity assessed (36: results per study, per tool), followed by criterion validity (31 results), convergent/divergent validity (14 results), internal consistency (7 results), structural validity (4 results) and reliability (3 results). Nine studies also specified that they had a "diagnostic accuracy" design (or similar) (Table 1.6).

After consideration of risk of bias, attention can be paid to the measurement properties of sufficient (+), insufficient (-) or indeterminate (?). Given the poor quality of many discriminant validity study designs, the measurement properties of these studies were rated as indeterminate (?). It should be noted that tools varied in language version, which may impact the confidence in results. Given the heterogeneity of the literature and the poor quality of reporting (e.g. in divergent validity studies) it was not possible to evaluate the psychometrics of tools across severities.

Structural Validity

Insufficient outcomes were reported from the Rasch analyses for the Cognistat to be evaluated using COSMIN. However, MMSE and MoCA (Swahili versions) both demonstrated sufficient structural validity in a TBI sample (Table 1.3).

Internal Consistency

The only measure which demonstrated sufficient internal consistency, with a low risk of bias rating, was the MoCA. The RUDAS and SLUMS appeared to have sufficient measurement properties, however, were rated as indeterminate, as information could not be found on their structural validity in this sample. The Cognistat and RQCST (Revised Quick Cognitive Screening Test) had insufficient internal consistency in particular subtests. There was insufficient information on the internal consistency of the MMSE (Table 1.4)

Reliability

A small number of studies found the RUDAS (Rowland Universal Dementia Assessment Scale) and Cognistat, had sufficient inter-rater reliability. One study found the RUDAS had sufficient test-re-test reliability. However other studies of test-rest reliability (RUDAS and Cognistat) suffered from sources of bias (time between testing) (Table 1.4).

Criterion Validity

Four studies found the MoCA had sufficient criterion validity to distinguish between TBI patients and controls. An additional study found the MoCA had sufficient criterion validity for those who were impaired/not impaired on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). Similarly, four studies found the MMSE had sufficient criterion validity to distinguish between TBI patients and controls. However one study found varying AUCs for MMSE domains to predict impairment on corresponding standard neuropsychological tests. The RUDAS, SLUMS (St Louis University Mental Status Examination), SCWT (Stroop Colour-Word Test), RQCST, COWAT TMT (Trail Making Test) and IFS (INECO Frontal Screen) all had initial promising evidence of their criterion validity, however were only supported by 1-2 studies. There were mixed findings for the SAC; and there is no evidence to support the criterion validity of the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT), Cognistat, CogState, BEAM, or King-Devick (K-D) tests (Table 1.5).

Hypothesis testing: Convergent validity

Given the heterogeneity of the measures used in convergent validity analyses it is difficult to draw conclusions across the studies. Cognitive screening tools, many of which have limited evidence of established validity in TBI samples, are mainly used as comparators. Studies did not often hypothesise the direction and strength of relationship expected, which is rated as indeterminate under COSMIN. In one study both the MMSE and MoCA were found to be moderately positively correlated with functional ability measures; and the RQCST was found to be correlated with activities of daily living and quality of life measures (Table 1.5).

Hypothesis testing: Discriminant validity

As discussed, given the lack of clear hypotheses and effect size reporting, it is difficult to draw meaningful conclusions from the divergent validity studies. Two studies found that there were significant differences in SAC scores between TBI and control groups (with large effect sizes). However, in one of these studies, the group means were not reported and in the other, the difference between groups was too small to be clinically useful. One study found a significant difference in MMSE score between TBI (mean total score = 25.00) and control groups (mean

total score =28.59) with a large effect size, suggesting those with TBI in the acute phase performed more poorly (Table 1.5).

Table 1.1 Sample Characteristics

ID	Reference	TBI group (n)	TBI Mean Age (years) (SD)	TBI % Male	TBI Severity (n)	TBI time since injury M (SD)	Control Group (n) (if present)	Controls Mean Age (years) (SD)	Controls % Male
1	(Adjorlolo, 2018)	50	36.28 (7.20)	70.00%	moderate	*8 months (4.58)*	50	27.98 (6.75)	58.00%
2	(Adjorlolo, 2016)	50	36.28 (7.20)	70.00%	moderate	8 months (4.58)	50	*27.98 (6.75)*	*58.00%*
3	(Borgaro and Prigatano, 2002)	42	Moderate= 34.09 (14.80); Severe = 38.05 (14.72)	moderate 77.3%; severe 66.7%	moderate (22); severe (20)	Moderate =16.05 days (11.91) severe=20.43 days (11.68)	21	37.67 (17.22)	76.20%
4	(Borgaro et al., 2003)	65	38.8 (16.02)	75.40%	mild (20), moderate (12); severe (33)	23.02 days (18.31)	25	38.63 (16.64)	59.30%
5	(Cheng et al., 2021)	86	51.651 (16.487)	70.90%	mild (78) moderate (4) severe (4)	*not reported*	40	53.250 (13.897)	70.00%
6	(Cheng et al., 2020)	50	51.92 (19.070)	78%	mild (43); Moderate (3) Severe (4)	8 days (7–11.25) *(median, IQR)*	32	50.88 (17.716)	68.80%
7	(Cole et al., 2018)	59 [1]	27.0 (6.0)	95.40%	mild	4.97 days (1.76)	68 [1]	34.4 (7.9)	79%
8	(Doninger et al., 2006)	120	37.5 (12.6)	82%	mild (30.6%), moderate (24.6%) severe (44.8%)	29 days (29)			
9	(Ettenhofer et al., 2021)	191	35.04 (8.05)	93.2	mild	27.12 months (8.28 - 85.81) *(median, IQR)*			

ID	Reference	TBI group (n)	TBI Mean Age (years) (SD)	TBI % Male	TBI Severity (n)	TBI time since injury M (SD)	Control Group (n) (if present)	Controls Mean Age (years) (SD)	Controls % Male
10	(Fischer et al., 2016)	11	33 (16.5)	72.72%	mild	5.9 hours (5.2)	12 HC; 7 O	O= 31 (11.6); HC= 33 (15.0)	O = 100%; HC= 75%
11	(Frenette et al., 2018)	134; 92 'complicated mTBI', 42 uncomplicated	Uncomplicated: 47.5 (26.0–59.0) Complicated: 58.5 (35.5–76.5)	Uncomplicated 66.67%; Complicated 60.87%	mild	"within the first 2 weeks following the injury"	25	21.92 (5.26)	48%
12	(Gupta and Kumar, 2009)	30	33.5 (10.7)	83.30%	moderate-severe	27.6 months (33.8)	55	26.53 (8.48).	47.27%
13	(Hatta et al., 2012)	42	30.7 (11.0)	80.95%%	"mostly mild TBI"	7.7 months (2.6)	32	50.88 (17.716)	68.80%
14	(Joseph et al., 2019)	100	43.47 (16.43);	overall 65%; mild 65%; moderate/severe 65%	Mild (60) moderate -severe (40)	386.14 days (454.49)			
15	(Tay et al., 2019)	61 (50 outpatient; 11 inpatient)	52.6 [17.4]	80.30%	Moderate- severe TBI	17.5 months (18.0)			
16	(Nabors et al., 1997)	45 [1]	39.5 (15.7)	78.00%	mild (21); moderate (8); severe (13); unknown (2)	34.7 days (25.3)			
17	(Pinasco et al., 2021)	28	36.59 (13.76)	71.43%	Mild (3); moderate-severe (25	60.33 months (61.02)			
18	(Rojas and Bennett, 1995)	25	27.44 (7.11)	52%	mild TBI				

ID	Reference	TBI group (n)	TBI Mean Age (years) (SD)	TBI % Male	TBI Severity (n)	TBI time since injury M (SD)	Control Group (n) (if present)	Controls Mean Age (years) (SD)	Controls % Male
19	(Silverberg et al., 2014)	26 [1]	36.6 (12.2)	73.10%	mild TBI		33	*4*2.8 (12.2)	51.5%
20	(Srivastava et al., 2006)	43 [1]	66 (7.2)	37.20%	mild (28); moderate (15)	"one year post-TBI"			
21	(Vissoci et al., 2019)	192	33.87 (13.32)	82.80%	Mild(175) Moderate -severe (17)				
22	(Waldron-Perrine et al., 2019)	117 [2]	31.33 (8.48)	95.7% (total sample)	mild TBI only				
23	(Walsh et al., 2016)	100	26.31 (5.83)	87%	mTBI only	≤ 72 h after injury	100	26.31 (5.83)	79%
24	(Wu et al., 2021)	98	51.17 (16.81)	75.51%	moderate-severe	20.23 days (5.16)	30	46.77 (16.13)	60%
25	(Zhang et al., 2016)	103	35.9 (13.08) [3]	84.47%		In the last "1-12 months"	42	31 (11.3)	80.95%%
26	(Zhang et al., 2021)	42	47.00 (15.814)	76.20%		*not reported*	30	44.07 (11.435)	60.00%
27	(Maruff et al., 2009)	50	43.2 (5.6)	80%	mTBI only	72 days (14)	50	44.7 (4.9)	80%
28	(Wong et al., 2013)	48	50 (16)	73%	Median (IQR) GCS score: 14 (11-15).	3–5 years	40		
29	(Luoto et al., 2014)	44	37.5 (12.1)	67.30%	mTBI only	38.4 hours (30.3)	33	42.8 (12.2)	51.50%
30	(Coldren et al., 2010)	71	26.5	96%	"concussive event" (moderate/severe excluded)	>12 hours after the event	102 HC and 64 I	27.3	88%

ID	Reference	TBI group (n)	TBI Mean Age (years) (SD)	TBI % Male	TBI Severity (n)	TBI time since injury M (SD)	Control Group (n) (if present)	Controls Mean Age (years) (SD)	Controls % Male
31	(Doninger et al., 2000)	186	34	75%		32 months *(Median)*			
32	(Stone et al., 2015)	84	36.1 (13)	76%	"concussion"	23.6 hrs (14)	30	36.9 (13)	76.7
33	(Bin Zahid et al., 2016)	118 (87 [-CT]; 31 [+CT])	41.09 (11.37) (-CT); 37.86 (13.91) (+CT)	76% (-CT); 81% (+CT)		0–5 days	98 HC; 46 O	HC = 32.14 (11.26); O = 35.72 (10.58)	HC =49% O= 80%

Abbreviations: TBI, Traumatic Brain Injury; mTBI, mild traumatic brain injury; CT+, TBI with "positive: CT scan; CT-, TBI with "negative" CT scan; HC, health controls; O, orthopedic injury; I, injured controls

1. *Maximum sample size (not all participants completed all screening tests)) 2. This total sample includes patients who were then confirmed NOT to have a diagnosis of TBI as part of this study; sample size of confirmed TBI and non-TBI group not reported. 3. Age at "onset" of TBI reported*

Table 1.2. Screening Tools

Screening Tool	Original Reference	Domains	Items	Range of Scores;	Time to administer	Included studies which evaluate the screen (study ID)
Axon Sport's CogState Sport (CogState)	(Collie et al., 2003)	psychomotor function/information processing, decision making, working memory and new learning	4		?	7, 27
Barrow Neurological Institute Screen for Higher Cerebral Functions (BNIS)	(Wass, 1997)	language functions, orientation, memory, concentration/attention, visual spatial problem-solving skills, affect expression and perception, and awareness			0-25 minutes [2]	3, 4
Bethesda Eye & Attention Measure (BEAM)	(Ettenhofer et al., 2016)	Visual attention			10 minutes	9
Cognistat (Neurobeh-avioral Cognitive Status Examination; NCSE)	(Kiernan et al., 1987)	Orientation, attention, language, constructional praxis, memory, calculations, and verbal reasoning. Level of conscious also reported.	51	average, mildly impaired, moderately impaired, or severely impaired.	10-20 minutes	8, 12, 16, 31
COntrolled Word Association Test (COWAT)	(Spreen and Benton, 1977)	Verbal fluency	3		3 minutes	1, 2
Digit cancellation test (D-CAT)	(Hatta et al., 2001)	Focused attention, sustained attention or concentration, and selective attention.			5 minutes	13
Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT)	(Maroon et al., 2000)	processing speed, reaction time, visual memory, and verbal memory			20 minutes	7
INECO Frontal Screening (IFS)	(Torralva et al., 2009)	executive function (motor programming, conflicting instructions, motor inhibitory control, verbal and		0-30	10 minutes	17

Screening Tool	Original Reference	Domains	Items	Range of Scores;	Time to administer	Included studies which evaluate the screen (study ID)
		visual working memory, verbal abstraction ability and inhibitory control)				
King-Devick Test (K-D)		cognitive processing speed and rapid gaze shifting			~2 minutes	10, 19, 23
Mini-Mental State Examination (MMSE)	(Folstein et al., 1975)	orientation, memory, calculation and attention, naming, language and visual space function.	11	0-30	5-10 mins	5, 6, 20, 21, 24, 25, 26, 28
Montreal Cognitive Assessment (MoCA)	(Nasreddine et al., 2005)	visuospatial and executive function, naming, memory, attention, language, abstraction, orientation	30	0-30	~10 mins	6, 11, 15, 21, 22, 24, 25, 26, 28
Revised Quick Cognitive Screening Test (RQCST)	(Mate-Kole et al., 2009)	Orientation, Attention/ concentration (verbal), Attention/concentration (visual), Memory: Immediate recall (verbal), Spatial neglect, Arithmetic, Constructional praxis, Memory: Immediate recall (visual), Vocabulary, Naming, Abstract reasoning: Similarities, Abstract reasoning: Analogies, Unusual views, Spatial orientation, Memory: Delayed recall (visual), Memory: Delayed recall and Memory: New learning.	48		10-15 mins	1, 2,
Rey's Rey Tangled Lines Test (RTLT)	(Rey, 1958)	Processing speed			< 5 minutes	14
Rowland Universal Dementia Assessment Scale (RUDAS)	(Storey et al., 2004)	memory, visual spatial orientation, visual structure painting, practical imitation, judgment, language	6	0-30	<10 minutes	5, 6,
Saint Louis university mental status examination (SLUMS)	(Zhang et al., 2021)	orientation, memory, calculation, attention, language, visual space and executive function	11	0-30	5-10 minutes	24, 26

Screening Tool	Original Reference	Domains	Items	Range of Scores;	Time to administer	Included studies which evaluate the screen (study ID)
Standard Assessment of Concussion (SAC)[1]	(McCrea et al., 1997)	orientation, immediate memory, concentration, and delayed recall.		0-30	5 minutes	10, 19, 29, 30, 32, 33
King-Devick Test (K-D)	(Oride et al., 1986)	cognitive processing speed and rapid gaze shifting			~2 minutes	
Stroop Colour Word Test (SCWT) / Stroop Neuropsychological Screening Test (SNST)	(Golden and Freshwater, 1978); (Trenerry et al., 1989)	Selective attention and interference control; "effectiveness of focused attention"			4 minutes	1, 2, 18
Trail Making Test (TMT)	(Reitan and Wolfson, 1995)	speed processing, sequence alternation, cognitive flexibility, visual search, motor performance, complex attention.	2			1, 2,

1. *The Standard Assessment of Concussion (SAC) was administered as part of the Military Assessment of Concussion (MACE) (French et al., 2008) or the Sport Concussion Assessment Tool (SCAT) (Guskiewicz et al., 2013); it was only reported if it was analysed separately (as both of these tools contain ineligible elements). 2. Estimates vary across the literature; as some are less than 20 minutes, this measure was included.*

Table 1.3 COSMIN: Structural Validity

Tool	Study ID	n	Type of Analysis	Results (rating)	RoB
Cognistat	8	Inpatient = 120; community = 296	Rasch	INPATIENT: 3 strata, person separation index 2.18 (reliability =.83) item separation index (3.71; reliability=.93), mistargeting logits (2.18), 5 items demonstrated misfit (mean square infit 1.30) COMMUNITY: 2 strata, 1.69 (reliability = .74) (3.73, corresponding to a reliability of .93), mistargeting logits (2.80) (?)	V/A
Cognistat	31	186	Rasch	person separation index =1.57 (reliability = .71); RMSE = 5.16; item separation index = 3.01, (reliability = .90). Recall of the word "green" and orientation to date demonstrated misfit (mean square infit statistic = 1.57 and 1.34, respectively) (?)	A
MMSE (Swahili)	21	192	Confirmatory Factor Analysis	RMSEA= 0.06 (95% CI= 0.04,0.08); CFI= 0.85, TLI=0.81. (+)	V
MoCA (Swahili)	21	192	Confirmatory Factor Analysis	RMSEA= 0.04 (95% CI=0.00-0.07); CFI=0.98; TLI=0.98. (+)	A

Abbreviations: MMSE, Mini Mental State Exam; MoCA, Montreal Cognitive Assessment; RoB=Risk of Bias (COSMIN): V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable. Measurement Property rating (COSMIN): (+) sufficient (+), insufficient (–), or indeterminate (?)

Table 1.4 Internal Consistency and Reliability

Tool	Study ID	Internal Consistency			Reliability (Inter-rater)				Reliability (Test Re-test)			
		n	Cronbach's **α** (rating)	RoB	n	Comparison	ICC (rating)	RoB	n	Comparison	ICC (rating)	RoB
Cognistat (Indian)	12	30	Total= .94, Subtests .61-.89 (-)	V	10	two independent scorers rated the performance from a video recording .	.82-1.00 (+)	V	8	Sub-sample of patients tested twice; the interval between testing ranged from 15 days to 30 days	.31-1.00 (-)	I
MMSE	21	192	.63 (-)	V								
MoCA	21	192	.78 (+)	V								
MoCA	28	48	.82 (+)	V								
RQCST	1	50	Verbal = .73 Nonverbal = .62 Global = .82 (-)	I								
RUDAS	6	50	.733 (?)	V	50	all three doctors scored the patient simultaneously	.933 (+)	V	50	Chief reviewer repeated tests 24-48hrs apart	.938 (+)	V
SLUMS	26	42	.723. (?)	I	42	three physicians, results were scored simultaneously	.983-.998 (+)	V	42	Within 24–48 h of the first evaluation, the main evaluator conducted another SLUMS assessment of the TBI patients again.	language, memory, visual space and executive function all ">0.75"; orientation, 0.740; Calculation and Attention, 0.645; (-)	D

Abbreviations: MMSE, Mini Mental State Exam; MoCA, Montreal Cognitive Assessment; RQCST Revised Quick Cognitive Screening Test, ; RUDAS, Rowland Universal Dementia Assessment Scale; SLUMS, St Louis Mental State Examination. RoB=Risk of Bias (COSMIN): V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable. Measurement Property rating: (+) sufficient (+), insufficient (–), or indeterminate (?)

Table 1.5 Criterion Validity and Hypotheses Testing

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
(Axon Sport) CogState	7	58 (max)	Pearson correlations; neuropsychology-ical test battery	r=.000-.406 (-)	V								
(Axon Sport) CogState (Hong Kong Version)	27									n=50, Detect 2.59 (.16), Identify 2.78 (.10), One-back .93 (.39), Learn .69 (.25)	n=50, Detect 2.46 (.06), Identify 2.69 (.09), One-back 1.22 (.12), Learn 1.06 (.15)	t tests. Detect 5.9, p<.0001; Identify 4.8, p<.0001; One Back 5.7, p<.0001; Learn 9.3, p<.0001 [2] (?)	D
BNIS	3									(n=22) 11.35 (15.20)	(n=50) 49.90 (12.7)	F= 36.02 (p<0.001) (?)	I
BNIS (orientation)	4									(n=65) 53.8% disoriented to time; 29.92% disorientated to place;	(n=25); 11.1% disoriented to time; 0% disorientated to place	Time ($\chi 2$ = 14.69, p < .01); Place ($\chi 2$ = 12.97, p < .01) [1] (?)	I
BEAM	9	?	Partial pearson correlations (age	r=.25 -.50 (-)	V								

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
			as covariate); BEAM composite X neuropsychological test battery										
Cognistat (Indian)	12					?	MMSE subtests (X Cognistat subtests)	.09 - .95 (?)	V	(n=30); orientation 9.47 (3.54), attention 5.47 (1.68) , comprehension 4.83 (1.51), repetition 10 (2.17) , naming 6.20 (2.11), immediate visual memory 1.20 (0.76), construction 2.13 (1.94) , memory 7.03 (3.66), calculation 3.20 (1.10), similarities 4.33, (2.12) and	(n=55) orientation 11.93 (.26) , attention 6.93 (1.21), comprehension 5.84 (.42) , repetition 11.45 (1.09), naming 7.75 (.55) , immediate visual memory 1.87 (.34), construction 4.91 (1.28), memory 11.07 (1.61) , calculation 3.84 (.37),	t-tests. orientation 5.41 (p<.001) , attention 4.61 (p<.001), comprehension 4.62 (p<.001) , repetition 4.13 (p<.001), naming 5.14 (p<.001) , immediate visual memory 5.64 (p<.001), construction 7.91 (p<.001), memory	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
										judgment 3.63 (1.59).	similarities 7.16 (1.15), and judgment 5.13 (1.02).	7.05 ($p<.001$) , calculation 3.92 (0.37), similarities 7.99 ($p<.001$), and judgment 5.27 ($p<.001$). (?)	
Cognist-at	16	?	Pearson zero order correlations; neuropsychological test battery	r=-.30-.68 (-)	V								
COWAT	1	100	ROC: TBI vs controls	AUC= .787 (+)	V					(n=50) 45.02 (6.90)	(n=50) 54.98 (10.19)	F(1,96)= 31.18, $p<0.01$ $\eta 2$= .25 (+)	V
COWAT	2					50	RQCST, SCWT, TMT, GADL, QOLIBRI	Cognitive screens: .10- .38, GADL: .15 QOLIBRI: .11 (-)	V				
DCAT	13									(n= 42) TOTAL	TBI (n= 42) TOTAL	significant main effect	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
										Trial 1=220.1 (79.1); 2=156.6 (45.5); 3 =127 (34.9) -	Trial 1=378.2 (88.4); Trail 2= 292.4 (47.8); Trial 3 = 224 (10.1)	of group on Total scores, F(1, 82) = 72.72, p < .001. (?)	
ImPACT	7	58 (max)	Pearson correlations; neuropsychological test battery	r=.000-0.475 (-)	V								
IFS	17		<1.5 SD below mean in at least 1 EF test	AUC= .95 (+)	V					n=28; total 21.96 (4.52),	n=32; total 27.90. (1.49)	Mann Whitney U = 37.500, p = .000 (?)	I
K-D	10	?	ROC: TBI vs controls	AUC=0.53 (-)	V					n= 7 ?	O, n=6; N, n=6 ?	p = .72(?)	I
K-D	19									(n=26) 51.50 (12.16)	(n=33) 47.09 (10.00)	F=2.16, p= 0.148, cohen's d= 0.40 (+)	V
K-D	23									(n=72) results in Mdn; IQR seconds Trial 1, 58.29	(n=32) results in Mdn; IQR seconds Trail 1, 44.93, (39.21–	Mann Whitney U tests. Trial 1, U = 2168, p ≤ 0.001; Trial 2,	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
										(49.41-72.97) Trial 2 53.49, (45.70–70.94)	50.49 s); Trail 2 42.80 (37.13–47.97)	U = 2380, p ≤ .001 (?)	
MMSE - Chinese	5	126	ROC: TBI vs controls	AUC= .721 (+)	V	86	RUDAS	r=.611(?)	V	(n=86) 24.977 (4.807)	(n=40) 28.125 (1.856)	"t-test" p< .001 (-)	I
MMSE - Chinese?	6	82	ROC: TBI vs controls	AUC= .769 (+)	V					(n=50) 25.00 (5.107)	(n=32) 28.59 (1.388)	cohen's d = 0.959, p<.001 (?)	V
MMSE	20	Varied by comparison	Sensitivity/ specificity; MMSE domains and impairment on corresponding standard neuropsychological tests (< 1.5 s.d. below the mean)	0-100 %/ 12.0-97.5% (-)	V								
MMSE Swahili	21					192	MoCA, FIM	MoCA, r = .68, FIM, r = .35 (+)	V				
MMSE	24	98	ROC curve: TBI vs controls	AUC= .8381(+)	V					n=98. 24.76 (4.70)	n=30, 28.97 (1.29)	t-test or wilcoxon rank sum	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
												test. p<.0001* (t-test or wilconxon signed rank) (?)	
MMSE Chinese	25	145	Sensitivity: TBI vs controls	Sensitivity: 69.90% (?)	I								
MMSE - Chinese?	26	72	ROC: TBI vs controls	AUC = .756 (+)	V					n=42, 25.64 (4.898)	n=30, 27.63 (2.798)	p=.009 (-)	I
MMSE - Cantonese	28									(n=48) Median, (IQR): 28 (26–30)	(n=40) Median, (IQR): 30 (28–30)	p< .001 (?)	I
MoCA - Chinese?	6	82	ROC: TBI vs controls	AUC=.824 (+)	V					TBI (n=50) 18.16 (6.600)	C (n=32) 25.25 (2.806)	cohen's d= 1.398) p<.001 (?)	V
MoCA - English or French	11									*Results in median (IQR).* u- TBI (n=42) 24.5 (22-27) c-TBI (n=42) 21(14.5-25)	*Results in median (IQR).* (n=25) 28 (27-29)	Total Chi2 = 68.1 p<.0001*; pairwise post hoc: c-mTBI < u- mTBI; c-mTBI < C; u-mTBI < C. (?)	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
MoCA - Singapo-re	15	?	ROC: MoCA vs RBANS (<5th centile) Sensitivity/specificity: CTT(<5th centile)	MoCA X RBANS AUC = 0.791. MoCA X CTT =79.4%/74.1% (+)	V/D								
MoCA - Swahili	21					192	MMSE, FIM	MMSE, r = .68; cFIM, r = .43 (+)	V				
MoCA	22									n= (?) 25.27 (2.12)	n= (?) 25.22 (2.51)	*t*(131) = −.12, *p* = .90 (?) **controlling for performanc-e validity*	I
MoCA	24	98	ROC curve: TBI vs controls	AUC= .8658 (+)	V					n=98, 18.60 (5.86)	n=30, 25.80 (1.88)	t-test or wilcoxon rank sum test. p<.0001* (-) (t-test or wilconxon signed rank)	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
MoCA - Bejing Version	25	145	Sensitivity: TBI vs controls	Sensitivity: 94.17% (?)	I								
MoCA - Chinese	26	72	ROC: TBI vs controls	AUC= .916 (+)	V					n=42, 19.48 (5.824)	n=30, 22.97 (4.723)	p=.009 (?)	I
MoCA Hong Kong Version	28	88	ROC: TBI vs controls	AUC= .704 (+)	V					*Results in Median, IQR* n=48, 24 (21–27)	*Results in Median, IQR* n=40: 26 (24–29)	p< .001 (?)	I
RQCST	1	100	ROC: TBI vs controls	AUC= .894 (+) (.674-.912 per subtest)	V	50	SCWT, COWAT, TMT,	.27-.40 (+)	V	(n=50) 44.07 (9.24)	(n=50): 55.93 (6.69)	f(1,96)=46.18, p<.001, η2=.33 (+)	V
RQCST	2					50	SCWT, COWAT, TMT (cognitive measures), GADL, QOLIBRI	RQCST (Global) X cognitive screens : .34 – .50, G-ADL: .72, QOLIBRI: .71 (+)	V				
RTLT	14									RTLT Latency: mTBI (n=60) (M = 9.66, SD = 4.79)		RTLT Latency t(98) = −1.67, p = .10.	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
										Moderate/severe (40) (M = 11.20, SD = 4.02) RTLT Accuracy: mTBI (M = 5.52, SD = .98) moderate/severe (M = 5.70, SD = .65)		RTLT accuracy t(98) = −1.04, p = .30. (?)	
RUDAS (Chinese)	5	126	ROC: TBI vs controls	AUC= .711 (+)	V	86	MMSE	.611 (?)	V	(n=86) 22.151 (4.565)	(n=40) 25.325 (2.596)	"t-test" p< .001 (?)	I
RUDAS Chinese	6	82	ROC: TBI vs controls	AUC= .844 (+)	V	50	MMSE, MoCA	MMSE: .701, MoCA; .778 (?)	V	(n=50) 21.00 (5.440)	(n=32) 26.41 (1.521)	cohen's d= 1.354, p<.001 (?)	V
SLUMS Chinese	24	98	ROC curve: TBI vs controls	AUC= .891 (+)	V	98	MMSE and MoCA	Spearman's rho MMSE= .76 MoCA =.84 (?)	V	n=98, 18.63 (5.99)	n=30, 26.40 (1.87)	t-test or wilcoxon rank sum test. p<.0001* (t-test or wilconxon	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
												signed rank) (?)	
SLUM-S Chinese	26	72	ROC: TBI vs controls	AUC= .872 (+)	V	42	MMSE and MoCA,	MMSE = .702, MoCA .831 (?)	V	n=42, 19.67 (5.498)	n=30, 23.77 (4.747)	p=.002 (?)	I
SAC	10	?	ROC: TBI vs controls	AUC= .84 (+)	V					(n = 7) ?	(n =6) ?	t[16] = −2.61, p = .019 (?)	I
SAC	19									(n=22) 26.00 (2.29)	(n=33) 27.67 (1.81)	F=9.025, p= .004, cohen's d= 0.81 (+)	V
SAC	29	77	ROC: TBI vs controls	AUC= .759 (+)	V	44	General Trauma Severity (ISS), trauma severity (divergent validity)	Spearman's rho= -.17 (?)	I	?	?	Mann-Whitney's U = 323; p < .001; d = 1.08 *TBI< controls* (+)	A
SAC	30	71 TBI; 168 controls	ROC:TBI vs controls	AUC=.5878 (-)	V					(n=71) 26	(n=168) 26.8	t-test; p=.02 (?)	I
SAC	32	84 concussion; 30 control	ROC; concussion vs controls	AUC=.650 (-)	V					n=84, 23.5 (4.8)	n=30, 25.9 (2.7)	p=.001 (?)	I
SAC	33	216	ROC: TBI (both + and -CT) vs HCs	AUC = .774 (+)	V					(-CT) (n=87) 22.0 (4.9) (+CT)	HC n=98, 26.0 (2.5);	Post-hoc bi-group comparison	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
										(n=31) 21.4 (4.6)		s; -CT vs HC p<.001; +CT vs HC p<.001 (p<.001 for significance, multiple comparisons) (?)	
SCWT	1	100	ROC: TBI vs controls	AUC= .793-.898 (+)	V					(n=50) CW 43.89 (7.34)	(n=50) CW 56.11 (8.48)	F(1,96)=46.15 , p<0.001 $\eta 2$= .33 (+)	V
SCWT	2					?	RQCST,TMT,COWAT, GADL, QOLIBRI	CW X cognitive screens: .17 -.45, GADL 0.43, QOLIBRI; 0.44 (-)	V				
SNST	18									(n=25); 82.92 (23.21)	(n=25) 97 (16.64)	Cut off score of 99; X2 (3, N = 50) = 2.96,	I

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
												p > .01. (accuracy .62) "not significant" (?)	
TMT	1	100	ROC: TBI vs controls	AUC= .746- .902 (+)	V					(n=50) TMT A =43.21 (8.19) ; TMTB = 44.97 (10.39);	(n=50) TMT A = 56.79 (6.41) ; TMTB = 55.03 (6.52)	TMT A, F(1,96)= 61.54 , p<.001 η2=.39 ; TMTB, F(1,96)=27.83 , p<.001 η2=.23 (+)	V
TMT	2					50	RQCST, SCWT, COWAT, GADL QOLIBRI	TMTA X cognitive screens: .10-.42, G-ADL: .30 QOLIBRI: .30 TMTB X cognitive screens: .30- .50,	V				

Tool	Study ID	Criterion Validity				Hypothesis Testing (Convergent unless specified)				Hypothesis Testing (Discriminant)			
		n	Comparison	AUC (or sensitivity/ specificity, or correlation) (rating)	RoB	n	Comparator	Pearson Correlation, (unless specified) (rating)	RoB	TBI Mean (SD)	Controls Mean (SD)	Result (rating)	RoB
								G-ADL: .16, QOLIBRI: .22 (-)					

Abbreviations: TBI: Traumatic Brain Injury; "C": controls; u-TBI, uncomplicated TBI; c-TBI, complicated TBI; mTBI, mild TBI; EF, executive function; AUC= area under the curve; ROC, ROC curve analysis. RoB, risk of bias, V very good, A adequate, D doubtful, I inadequate. Measurement Property rating: (+) sufficient (+), insufficient (–), or indeterminate (?).

Cognitive Screens: BNIS, Barrow Neurological Institute Screen for Higher Cerebral Functions; BEAM, Bethesda Eye & Attention Measure; Cogstate, Axon Sport's CogState Sport; Cognistat (Neurobehavioral Cognitive Status Examination; NCSE); COWAT, COntrolled Word Association Test; D-CAT, Digit cancellation test; ImPACT, Immediate Post-Concussion Assessment and Cognitive Testing; IFS, INECO Frontal Screening; K-D, King-Devick Test MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment, RQCST, Revised Quick Cognitive Screening Test; RTLT, Rey's Rey Tangled Lines Test ; RUDAS, Rowland Universal Dementia Assessment Scale; SAC, Standard Assessment of Concussion; SCWT Stroop Colour Word Test; SLUMS, Saint Louis university mental status examination; SNST, Stroop Neuropsychological Screening Test; TMT Trail Making Test

Other measures: GADL Global Activities of Daily Living; QOLIBRI - quality of life measure; CTT colour trails test (REF); Functional Independence Measure (FIM); RBANS, The Repeatable Battery for the Assessment of Neuropsychological Status; ISS general trauma severity. (FIM); RBANS, The Repeatable Battery for the Assessment of Neuropsychological Status; ISS general trauma severity.

Table 1.5 cont. 1.Spinal injury group also included in chi squared analysis 2. Mean difference and confidence intervals reported, but difficult to accurately interpret (graphical format)

Table 1.6 QUADAS 2 Risk of Bias and Applicability Ratings (Diagnostic Test Accuracy Studies only)

Study ID	Risk of Bias				Applicability		
	patient selection	index test	Reference Standard	flow and timing	patient selection	index test	Reference Standard
1	HIGH	HIGH	LOW	UNCLEAR	LOW	LOW	LOW
5	HIGH	HIGH	UNCLEAR	UNCLEAR	UNCLEAR	LOW	LOW
6	HIGH	HIGH	UNCLEAR	UNCLEAR	LOW	LOW	LOW
15	UNCLEAR	HIGH	UNCLEAR	LOW	LOW	LOW	UNCLEAR
19	HIGH	HIGH	LOW	HIGH	LOW	UNCLEAR	LOW
24	HIGH	HIGH	UNCLEAR	UNCLEAR	LOW	LOW	LOW
28	HIGH	HIGH	UNCLEAR	UNCLEAR	LOW	LOW	LOW
29	HIGH	HIGH	LOW	UNCLEAR	UNCLEAR	LOW	LOW
32	LOW	HIGH	LOW	LOW	LOW	LOW	LOW

"HIGH" = risk of bias; "LOW" = low/no identified risk of bias; "UNCLEAR" when insufficient data to make a judgement (Whiting et al., 2011)

Discussion

This systematic review found a substantial literature on the psychometric properties of cognitive screening tools in TBI populations. The MoCA in particular, demonstrated sufficient internal consistency, structural and criterion validity. The MMSE demonstrated sufficient structural validity, discriminant and criterion validity. In contrast, criterion validity for the MMSE varied depending on the standard neuropsychological test comparator and evidence for its internal consistency across all items was insufficient. Tools such as RUDAS, SLUMS, SCWT, RQCST, COWAT, TMT and IFS demonstrated promising validity, but were limited to a small number of studies. The properties of other tools were either inconsistent, insufficient or indeterminate. These finding have implications for the refinement of clinical guidelines and should inform clinical practice. With the current evidence-base, the MoCA can be tentatively recommended as the most well-validated tool and should be highlighted to clinicians as the preferred cognitive screening tool for TBI populations (while also making clinicians aware of the limitations of the literature) (NICE, 2014; Scottish Acquired Brain Injury Network, 2017; Teager et al., 2020).

These findings should be considered within the context of the heterogeneity and poor reporting across studies in terms of key sample demographics (TBI severity and time since injury). Many studies assessed the psychometric properties in acute or mild TBI settings; and these findings are unlikely to be generalisable to other clinical contexts, such as post-acute rehabilitation settings (Barman et al., 2016). The heterogeneity in psychometric qualities assessed and high risk of bias in certain study designs (particularly discriminative validity), meant that it was not possible to synthesise evidence meaningfully across studies (for example, using a GRADE approach). Poor statistical reporting, poor reporting of TBI diagnosis methods, insufficient analyses for multidimensionality, and time delays between reliability measurements were just some of the risk of bias issues identified. Finally, while tools such as the MMSE did not demonstrate internal consistency across domains, this may not necessarily be interpreted as an indicator of poor psychometric qualities. High internal consistency across domains may not be expected for multi-dimensional tools. This is especially pertinent for TBI populations who may not experience general impairment affecting all domains and who instead may have impairments in discrete domains.

These findings are broadly consistent with the findings in ABI and stroke populations, which find that multi-dimensional tools which assess executive function, including the MoCA, are typically the most well-validated (Stolwyk et al., 2014; Kosgallana et al., 2019). TBI often results in multi-domain cognitive difficulties, including executive function; which may explain this finding (Barman et al., 2016). It is notable that other multi-domain tools such as the ACE-III or Oxford Cognitive Screen (OCS) (which has been identified to have good criterion validity in stroke samples) have not been validated in TBI (Hsieh et al., 2013; Demeyere et al., 2015; Kosgallana et al., 2019). Notably, many of the screening tools investigated, including the MoCA and MMSE, were originally designed for detecting cognitive impairment associated with dementia and subsequently, tend to feature less executive function items (Folstein et al., 1975; Nasreddine et al., 2005). An alternative approach may be to design a new, bespoke measure designed for use in TBI populations, which is tailored to capture the executive difficulties which are common in these patients (Barman et al., 2016).

A possible limitation is the exclusion of SRC samples, as those with SRC do present to clinical settings. However, the SRC literature had key distinctive features (noted in the methods section) which may have increased the heterogeneity of the studies further. These studies may be the subject of a subsequent review on this distinct literature. Using COSMIN criteria may have excluded studies which contained relevant psychometric information. However, it would have been impractical to screen every study which inadvertently measured a tool's validity. It also would not have been feasible or meaningful to attempt to rate the measurement properties of studies which used unusual designs or statistics which did not allow for comparison with COSMIN. Finally, applying age criteria resulted in several studies being excluded. However this decision was made as there are differences in the nature and assessment of TBI in adolescents or older adults, due to the interaction of developmental factors (Peters and Gardner, 2018; Christensen et al., 2021). A strength of this review is the use of two rigorous quality assessment tools, the COSMIN and the QUADAS 2, which contain detailed considerations on the assessment of psychometric methodology, and which have been specifically designed for use in these types of studies.

In conclusion, this review tentatively recommends the use of the MoCA in TBI populations, with the following caveats: 1) it's validity across key TBI demographics requires further clarification; 2) additional psychometric qualities including reliability and measurement error are evaluated. Tools such as RUDAS, SLUMS and RQCST show promise but require

additional investigation. The MMSE cannot be recommended currently due to insufficient evidence of its criterion validity. Future research should clarify the validity of tools identified in this review, and additional multidimensional tools which have been validated in similar populations or which are used in clinical practice, particularly the ACE-III and OCS. Psychometric studies should take consideration of COSMIN and QUADAS 2 guidelines in their design, to allow for meaningful comparison across tools.

References

Abd Razak, M. A., Ahmad, N. A., Chan, Y. Y., Mohamad Kasim, N., Yusof, M., Abdul Ghani, M. K. A., et al. (2019). Validity of screening tools for dementia and mild cognitive impairment among the elderly in primary health care: a systematic review. *Public Health* 169, 84–92. doi: 10.1016/j.puhe.2019.01.001.

Adjorlolo, S. (2016). Ecological validity of executive function tests in moderate traumatic brain injury in Ghana. *Clin Neuropsychol* 30, 1517–1537. doi: 10.1080/13854046.2016.1172667.

Adjorlolo, S. (2018). Diagnostic Accuracy, Sensitivity, and Specificity of Executive Function Tests in Moderate Traumatic Brain Injury in Ghana. *Assessment* 25, 498–512. doi: 10.1177/1073191116646445.

Barman, A., Chatterjee, A., and Bhide, R. (2016). Cognitive Impairment and Rehabilitation Strategies After Traumatic Brain Injury. *Indian Journal of Psychological Medicine* 38, 172–181. doi: 10.4103/0253-7176.183086.

Bin Zahid, A., Hubbard, M., Dammavalam, V., Balser, D., Pierre, G., Kim, A., et al. (2016). Assessment of acute head injury in an emergency department population using sport concussion assessment tool - 3rd edition. *Applied neuropsychology. Adult* 25, 1–10. doi: 10.1080/23279095.2016.1248765.

Borgaro, S. R., Kwasnica, C., Cutter, N., and Alcott, S. (2003). The use of the BNI screen for higher cerebral functions in assessing disorientation after traumatic brain injury. *J Head Trauma Rehabil* 18, 284–291. doi: 10.1097/00001199-200305000-00006.

Borgaro, S. R., and Prigatano, G. P. (2002). Early cognitive and affective sequelae of traumatic brain injury: a study using the BNI Screen for Higher Cerebral Functions. *J Head Trauma Rehabil* 17, 526–534. doi: 10.1097/00001199-200212000-00004.

Canadian Agency for Drugs and Technologies in Health (CADTH) (2014). Screening Tools to Identify Adults with Cognitive Impairment Associated with a Cerebrovascular Accident or Traumatic Brain Injury: Diagnostic Accuracy. Available at: https://www.cadth.ca/sites/default/files/pdf/htis/nov-2014/RB0751%20Cognitive%20Assessments%20CVA%20and%20TBI%20Final.pdf .

Cheng, Y., Wang, Y.-Z., Zhang, Y., Wang, Y., Xie, F., Zhang, Y., et al. (2021). Comparative analysis of rowland universal dementia assessment scale and mini-mental state examination in cognitive assessment of traumatic brain injury patients. *NeuroRehabilitation* 49, 39–46. doi: 10.3233/NRE-210044.

Cheng, Y., Zhang, Y., Zhang, Y., Wu, Y.-H., and Zhang, S. (2020). Reliability and validity of the Rowland Universal Dementia Assessment Scale for patients with traumatic brain injury. *Appl Neuropsychol Adult*, 1–7. doi: 10.1080/23279095.2020.1856850.

Christensen, J., Eyolfson, E., Salberg, S., and Mychasiuk, R. (2021). Traumatic brain injury in adolescence: A review of the neurobiological and behavioural underpinnings and outcomes. *Developmental Review* 59, 100943. doi: 10.1016/j.dr.2020.100943.

Coldren, R. L., Kelly, M. P., Parish, R. V., Dretsch, M., and Russell, M. L. (2010). Evaluation of the Military Acute Concussion Evaluation for use in combat operations more than 12 hours after injury. *Mil Med* 175, 477–481. doi: 10.7205/milmed-d-09-00258.

Cole, W. R., Arrieux, J. P., Ivins, B. J., Schwab, K. A., and Qashu, F. M. (2018). A Comparison of Four Computerized Neurocognitive Assessment Tools to a Traditional Neuropsychological Test Battery in Service Members with and without Mild Traumatic Brain Injury. *Arch Clin Neuropsychol* 33, 102–119. doi: 10.1093/arclin/acx036.

Collie, A., Maruff, P., Makdissi, M., McCrory, P., McStephen, M., and Darby, D. (2003). CogSport: Reliability and Correlation with Conventional Cognitive Tests Used in Postconcussion Medical Evaluations. *Clinical Journal of Sport Medicine* 13, 28–32.

Cullen, B., O'Neill, B., Evans, J. J., Coen, R. F., and Lawlor, B. A. (2007). A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery & Psychiatry* 78, 790–799. doi: 10.1136/jnnp.2006.095414.

Demeyere, N., Riddoch, M. J., Slavkova, E. D., Bickerton, W.-L., and Humphreys, G. W. (2015). The Oxford Cognitive Screen (OCS): Validation of a stroke-specific short cognitive screening tool. *Psychological Assessment* 27, 883–894. doi: 10.1037/pas0000082.

Doninger, N. A., Bode, R. K., Heinemann, A. W., and Ambrose, C. (2000). Rating Scale Analysis of the Neurobehavioral Cognitive Status Examination. *The Journal of Head Trauma Rehabilitation* 15, 683–695.

Doninger, N. A., Ehde, D. M., Bode, R. K., Knight, K., Bombardier, C. H., and Heinemann, A. W. (2006). Measurement properties of the neurobehavioral cognitive status examination (Cognistat) in traumatic brain injury rehabilitation. *Rehabilitation Psychology* 51, 281–288. doi: 10.1037/0090-5550.51.4.281.

Ettenhofer, M. L., Hershaw, J. N., and Barry, D. M. (2016). Multimodal assessment of visual attention using the Bethesda Eye & Attention Measure (BEAM). *Journal of Clinical and Experimental Neuropsychology* 38, 96–110. doi: 10.1080/13803395.2015.1089978.

Ettenhofer, M. L., Hungerford, L. D., and Agtarap, S. (2021). Multimodal Neurocognitive Screening of Military Personnel With a History of Mild Traumatic Brain Injury Using the Bethesda Eye & Attention Measure. *J Head Trauma Rehabil* 36, 447–455. doi: 10.1097/HTR.0000000000000683.

Fischer, T. D., Red, S. D., Chuang, A. Z., Jones, E. B., McCarthy, J. J., Patel, S. S., et al. (2016). Detection of Subtle Cognitive Changes after mTBI Using a Novel Tablet-Based Task. *J Neurotrauma* 33, 1237–1246. doi: 10.1089/neu.2015.3990.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6.

French, L., Mccrea, M., and Baggett, M. (2008). The Military Acute Concussion Evaluation (MACE). *Journal of Special Operations Medicine* 8, 68–77.

Frenette, L. C., Tinawi, S., Correa, J. A., Alturki, A. Y., LeBlanc, J., Feyz, M., et al. (2018). Early detection of cognitive impairments with the Montreal Cognitive Assessment in patients with uncomplicated and complicated mild traumatic brain injury. *Brain Inj*, 1–9. doi: 10.1080/02699052.2018.1542506.

Gaber, T. A.-Z. K. (2008). Evaluation of the Addenbrooke’s Cognitive Examination’s validity in a brain injury rehabilitation setting. *Brain Injury* 22, 589–593. doi: 10.1080/02699050802132461.

Golden, C. J., and Freshwater, S. M. (1978). *Stroop color and word test (SCWT): A Manual for Clinical and Experimental Uses*. Wood Dale, Illinois: Stoelting Company.

Gupta, A., and Kumar, N. K. (2009). Indian adaptation of the Cognistat: Psychometric properties of a cognitive screening tool for patients of traumatic brain injury. *The Indian Journal of Neurotrauma* 6, 123–132. doi: 10.1016/S0973-0508(09)80006-3.

Guskiewicz, K. M., Register-Mihalik, J., McCrory, P., McCrea, M., Johnston, K., Makdissi, M., et al. (2013). Evidence-based approach to revising the SCAT2: introducing the SCAT3. *Br J Sports Med* 47, 289–293. doi: 10.1136/bjsports-2013-092225.

Hatta, T., Itto, Y., and Yoshizaki, K. (2001). *D-CAT: Screening Test fo Attention*. Osaka: Union Press.

Hatta, T., Yoshizaki, K., Ito, Y., Mase, M., and Kabasawa, H. (2012). Reliability and Validity of the Digit Cancellation TEST, a brief screen of attention. *An International Journal of Psychological Sciences* 55, 246–256. doi: 10.2117/psysoc.2012.246.

Hazan, E., Zhang, J., Brenkel, M., Shulman, K., and Feinstein, A. (2017). Getting clocked: screening for TBI-related cognitive impairment with the clock drawing test. *Brain Inj* 31, 1501–1506. doi: 10.1080/02699052.2017.1376763.

Headway (2017). ABI 2016-2017 Statistics based on UK admissions. Available at: https://www.headway.org.uk/about-brain-injury/further-information/statistics/ [Accessed May 2, 2022].

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., and Hodges, J. R. (2013). Validation of the Addenbrooke’s Cognitive Examination III in Frontotemporal Dementia and Alzheimer’s Disease. *DEM* 36, 242–250. doi: 10.1159/000351671.

Joseph, A.-L. C., Peterson, H. A., Garcia, K. M., McNally, S. M., Mburu, T. K., Lippa, S. M., et al. (2019). Rey’s Tangled Line Test: A measure of processing speed in TBI. *Rehabil Psychol* 64, 445–452. doi: 10.1037/rep0000284.

Kosgallana, A., Cordato, D., Chan, D. K. Y., and Yong, J. (2019). Use of Cognitive Screening Tools to Detect Cognitive Impairment After an Ischaemic Stroke: a

Systematic Review. *SN Compr. Clin. Med.* 1, 255–262. doi: 10.1007/s42399-018-0035-2.

Lim, P. A., McLean, A. M., Kilpatrick, C., DeForge, D., Iverson, G. L., and Silverberg, N. D. (2016). Temporal stability and responsiveness of the Montreal Cognitive Assessment following acquired brain injury. *Brain Injury* 30, 29–35. doi: 10.3109/02699052.2015.1079732.

Luoto, T. M., Silverberg, N. D., Kataja, A., Brander, A., Tenovuo, O., Ohman, J., et al. (2014). Sport concussion assessment tool 2 in a civilian trauma sample with mild traumatic brain injury. *J Neurotrauma* 31, 728–738. doi: 10.1089/neu.2013.3174.

Maroon, J. C., Lovell, M. R., Norwig, J., Podell, K., Powell, J. W., and Hartl, R. (2000). Cerebral Concussion in Athletes: Evaluation and Neuropsychological Testing. *Neurosurgery* 47, 659–672.

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., et al. (2009). Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Arch Clin Neuropsychol* 24, 165–178. doi: 10.1093/arclin/acp010.

Mate-Kole, C. C., Conway, J., Catayong, K., Bieu, R., Sackey, N. A., Wood, R., et al. (2009). Validation of the revised quick cognitive screening test. *Arch Phys Med Rehabil* 90, 1469–1477. doi: 10.1016/j.apmr.2009.02.007.

McCrea, M., Kelly, J. P., Kluge, J., Ackley, B., and Randolph, C. (1997). Standardized Assessment of Concussion in football players. *Neurology* 48, 586–588. doi: 10.1212/WNL.48.3.586.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 19, 539–549. doi: 10.1007/s11136-010-9606-8.

Nabors, N. A., Millis, S. R., and Rosenthal, M. (1997). Use of the Neurobehavioral Cognitive Status Examination (Cognistat) in traumatic brain injury. *The Journal of Head Trauma Rehabilitation* 12, 79–84. doi: 10.1097/00001199-199706000-00008.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x.

NICE (2014). Head injury: assessment and early management. Available at: https://www.nice.org.uk/guidance/cg176/chapter/introduction [Accessed May 5, 2022].

Oride, M. K., Marutani, J. K., Rouse, M. W., and DeLand, P. N. (1986). Reliability study of the Pierce and King-Devick saccade tests. *Am J Optom Physiol Opt* 63, 419–424. doi: 10.1097/00006324-198606000-00005.

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. *Systematic Reviews* 10, 89. doi: 10.1186/s13643-021-01626-4.

Peters, M. E., and Gardner, R. C. (2018). Traumatic brain injury in older adults: do we need a different approach? *Concussion* 3, CNC56. doi: 10.2217/cnc-2018-0001.

Pinasco, C., Oviedo, M., Goldfeder, M., Bruno, D., Lischinsky, A., Torralva, T., et al. (2021). Sensitivity and specificity of the INECO frontal screening (IFS) in the detection of patients with traumatic brain injury presenting executive deficits. *Appl Neuropsychol Adult*, 1–8. doi: 10.1080/23279095.2021.1937170.

Reitan, R. M., and Wolfson, D. (1995). Category test and trail making test as measures of frontal lobe functions. *The Clinical Neuropsychologist* 9, 50–56. doi: 10.1080/13854049508402057.

Rey, A. (1958). *L'examen clinique en psychologie. [The clinical examination in psychology.]*. Oxford, England: Presses Universitaries De France.

Rojas, D. C., and Bennett, T. L. (1995). Single versus composite score discriminative validity with the Halstead–Reitan Battery and the Stroop Test in mild brain injury. *Archives of Clinical Neuropsychology* 10, 101–110. doi: 10.1093/arclin/10.2.101.

Scottish Aquired Brain Injury Network (2017). Assessment of Cognition – Brain Injury E-Learning Resource. Available at: https://www.acquiredbraininjury-education.scot.nhs.uk/impact-of-abi/cognitive-problems/assessment-of-cognition/ [Accessed December 13, 2020].

Silverberg, N. D., Luoto, T. M., Öhman, J., and Iverson, G. L. (2014). Assessment of mild traumatic brain injury with the King-Devick Test in an emergency department sample. *Brain Inj* 28, 1590–1593. doi: 10.3109/02699052.2014.943287.

Spreen, O., and Benton, A. L. (1977). *Neurosensory Center Comprehensive Examination for Aphasia.* Victoria BC: University of Victoria Neuropsychology Laboratory.

Srivastava, A., Rapoport, M. J., Leach, L., Phillips, A., Shammi, P., and Feinstein, A. (2006). The utility of the mini-mental status exam in older adults with traumatic brain injury. *Brain Inj* 20, 1377–1382. doi: 10.1080/02699050601111385.

Stolwyk, R. J., O'Neill Megan H., McKay Adam J.D., and Wong Dana K. (2014). Are Cognitive Screening Tools Sensitive and Specific Enough for Use After Stroke? *Stroke* 45, 3129–3134. doi: 10.1161/STROKEAHA.114.004232.

Stone, M. E. J., Safadjou, S., Farber, B., Velazco, N., Man, J., Reddy, S. H., et al. (2015). Utility of the Military Acute Concussion Evaluation as a screening tool for mild traumatic brain injury in a civilian trauma population. *Journal of Trauma and Acute Care Surgery* 79, 147–151. doi: 10.1097/TA.0000000000000679.

Storey, J. E., Rowland, J. T. J., Conforti, D. A., and Dickson, H. G. (2004). The Rowland Universal Dementia Assessment Scale (RUDAS): a multicultural cognitive assessment scale. *International Psychogeriatrics* 16, 13–31. doi: 10.1017/S1041610204000043.

Tay, M. R. J., Soh, Y. M., Plunkett, T. K., Ong, P. L., Huang, W., and Kong, K. H. (2019). The Validity of the Montreal Cognitive Assessment for Moderate to Severe Traumatic Brain Injury Patients: A Pilot Study. *American Journal of Physical Medicine & Rehabilitation* 98, 971–975. doi: 10.1097/PHM.0000000000001227.

Teager, A., Methley, A., Dawson, B., and Wilson, H. (2020). The use of cognitive screens within major trauma centres in England: A survey of current practice. *Trauma* 22, 201–207. doi: 10.1177/1460408619871801.

Torralva, T., Roca, M., Gleichgerrcht, E., López, P., and Manes, F. (2009). INECO Frontal Screening (IFS): A brief, sensitive, and specific tool to assess executive functions in dementia–CORRECTED VERSION. *J Int Neuropsychol Soc* 15, 777–786. doi: 10.1017/S1355617709990415.

Trenerry, M., Crosson, B., DeBoe, J., and Leber, W. (1989). *Stroop Neuropsychological Screening Test (SNST) | PAR*. Odessa, FL: Psychological Assessment Resources (PAR) Available at: https://www.parinc.com/Products/Pkey/421 [Accessed July 16, 2022].

Vissoci, J. R. N., de Oliveira, L. P., Gafaar, T., Haglund, M. M., Mvungi, M., Mmbaga, B. T., et al. (2019). Cross-cultural adaptation and psychometric properties of the MMSE and MoCA questionnaires in Tanzanian Swahili for a traumatic brain injury population. *BMC Neurology* 19, 57. doi: 10.1186/s12883-019-1283-9.

Waldron-Perrine, B., Gabel, N. M., Seagly, K., Kraal, A. Z., Pangilinan, P., Spencer, R. J., et al. (2019). Montreal Cognitive Assessment as a screening tool: Influence of performance and symptom validity. *Neurol Clin Pract* 9, 101–108. doi: 10.1212/CPJ.0000000000000604.

Walsh, D. V., Capó-Aponte, J. E., Beltran, T., Cole, W. R., Ballard, A., and Dumayas, J. Y. (2016). Assessment of the King-Devick® (KD) test for screening acute mTBI/concussion in warfighters. *Journal of the Neurological Sciences* 370, 305–309. doi: 10.1016/j.jns.2016.09.014.

Wass, P. J. (1997). *An Analysis of the Construct-Related and Ecological Validity of the Barrow Neurological Institute Screen for Higher Cerebral Functions.* Ontario, Canada: University of Windsor.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., et al. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine* 155, 529–536.

Wong, G. K. C., Ngai, K., Lam, S. W., Wong, A., Mok, V., and Poon, W. S. (2013). Validity of the Montreal Cognitive Assessment for traumatic brain injury patients with intracranial haemorrhage. *Brain Injury* 27, 394–398. doi: 10.3109/02699052.2012.750746.

Wu, Y., Wang, Y., Zhang, Y., Yuan, X., and Gao, X. (2021). A preliminary study of the Saint Louis university mental status examination (SLUMS) for the assessment of cognition in moderate to severe traumatic brain injury patients. *Appl Neuropsychol Adult*, 1–5. doi: 10.1080/23279095.2021.1952414.

Zhang, H., Zhang, X.-N., Zhang, H.-L., Huang, L., Chi, Q.-Q., Zhang, X., et al. (2016). Differences in cognitive profiles between traumatic brain injury and stroke: A comparison of the Montreal Cognitive Assessment and Mini-Mental State Examination. *Chinese Journal of Traumatology* 19, 271–274. doi: 10.1016/j.cjtee.2015.03.007.

Zhang, S., Wu, Y.-H., Zhang, Y., Zhang, Y., and Cheng, Y. (2021). Preliminary study of the validity and reliability of the Chinese version of the Saint Louis University Mental Status Examination (SLUMS) in detecting cognitive impairment in patients with traumatic brain injury. *Appl Neuropsychol Adult* 28, 633–640. doi: 10.1080/23279095.2019.1680986.

Chapter 2: Major Research Project

Validity of the Addenbrookes Cognitive Exam-III as a Cognitive Screening Tool After Acquired Brain Injury

Authors:

Jessica McLaren[1*], Alex Fradera[2], Dr Breda Cullen[3]

1. Mental Health and Wellbeing, Institute of Health and Wellbeing, University of Glasgow, Scotland, United Kingdom; and NHS Ayrshire and Arran, Scotland, United Kingdom
2. Mental Health and Wellbeing, Institute of Health and Wellbeing, University of Glasgow, Scotland, United Kingdom
3. Mental Health and Wellbeing, Institute of Health and Wellbeing, University of Glasgow, Scotland, United Kingdom;and NHS Lanarkshire, Scotland, United Kingdom

Corresponding author: j.mclaren.2@research.gla.ac.uk

Prepared in accordance with Frontiers in Psychology: Neuropsychology

https://www.frontiersin.org/journals/psychology/for-authors/author-guidelines

Word Count: 8183

Plain Language Summary

Title: Validity of the ACE-III as a Cognitive Screening Tool After Acquired Brain Injury.

Background: Many people who have an acquired brain injury (ABI) (damage to the brain after birth) experience cognitive impairment, such as difficulties with attention, memory or doing tasks which involve visual information etc. If we can detect when someone is experiencing cognitive impairment, they can get the right assessment and help. Cognitive screening tools, such as the Addenbrooke's Cognitive Examination (ACE-III) are quick tests which indicate if someone might have cognitive impairment. However, no research study has checked whether the ACE-III is valid for use in patients with ABI (i.e. whether it measures what it claims to measure).

Aims and Questions: This project investigated the relationship between the patient's score on the different sub-tests of the ACE-III (each of which measures a different aspect of cognitive function e.g. memory) and their score on a more thorough cognitive test. The following key research questions were addressed:

1. Is the ACE-III a valid screening tool for use in ABI?
2. How effective is the ACE-III at predicting whether a person with ABI has a cognitive impairment?

Methods: A research database was developed using existing routine clinical data from a brain injury rehabilitation centre. Participants had a diagnosis of ABI. They gave consent for their data to be used for service evaluation; and ethical approval, which took that context into account, was received. Demographic information, their scores on the ACE-III, and additional cognitive tests were collected, and transferred onto an anonymised research database to protect patients identities. Each patient's ACE-III scores were compared with their scores on a number of different cognitive tests, measuring the same aspects of cognition.

Main Findings: Patients' scores on the ACE-III and corresponding cognitive tests were correlated with each other for each area of cognitive impairment (including attention, language and visuospatial skills). This provides evidence of the ACE-III's validity, as it tells us that it appears to measure what it claims to measure (in these areas). A patient's ACE-III

visuospatial subtest score was also able to classify those with and without impairment on a standard cognitive test of visuospatial skills. While these results are promising, because there was a small number of participants and missing data, this meant that it could not be determined if some of the results were significant; and the estimates of correlations were not very precise.

Conclusions: This study found promising initial evidence for the use of the ACE-III cognitive screening tool. The findings suggest that the ACE-III can be used to detect visuospatial impairment, but additional research with larger sample sizes is needed to determine if it is a valid measure of other types of cognitive function.

References:

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J. R. (2013). Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal Dementia and Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *36*(3–4), 242–250. https://doi.org/10.1159/000351671

Word count: 488

Abstract

The Addenbrooke's Cognitive Examination (ACE-III) is a widely used cognitive screening measure. Yet, despite its face validity and clinical utility in acquired brain injury (ABI), it has not yet been formally validated for use in this population. This project aimed to validate the domains of the ACE-III (attention, memory, fluency, language and visuospatial functioning) against standard cognitive tests, in an ABI population. Using a quantitative, cross-sectional design, routine data were analysed from 26 patients with ABI admitted to a specialist neuro-behavioural assessment and rehabilitation centre. Patients completed the ACE-III, a battery of cognitive tests and a functional impairment measure, the Mayo Portland Adaptability Index (MPAI-4) before or shortly after admission. Scores on each ACE-III domain were correlated against corresponding standard cognitive test(s) and the MPAI-4. Receiver operating characteristic (ROC) curves were used to determine the ability of each ACE-III domain score to classify cognitive impairment. Moderate-strong correlations were found between ACE-III attention, language and visuospatial domains, with corresponding cognitive tests. Patients' scores on several ACE-III domains were correlated with functional outcome. Patients scores on the ACE-III visuospatial domain were able to classify those with/without visuospatial impairment. Data availability and missingness impacted the available power to detect weak-moderate correlations and satisfactory AUCs. This is the first study to provide preliminary evidence for the validity of the ACE-III in ABI. Further investigation of the ACE-III's validity and reliability is required in ABI populations.

Word Count: 231

Keywords: Acquired Brain Injury; ACE-III; Cognitive Screening; Neuropsychological Test; Validity; Psychometric Assessment

Introduction

Acquired brain injuries (ABI), which are brain injuries after birth, are a frequent cause of hospital admission in the UK. One patient is admitted to hospital with an ABI every 90 seconds (Headway, 2017). Cognitive impairments are common in the period immediately after ABI, and in mild cases often resolve within 3 months (Barman, Chatterjee and Bhide, 2016). However, these cognitive difficulties can become more persisting, especially in moderate-severe cases. For example, between 32-56% of stroke survivors continue to meet criteria for cognitive impairment three months after their stroke (Whyte *et al.*, 2011). Those with moderate or severe traumatic brain injury (TBI) typically experience "marked" cognitive impairment following their injury, which does not return to baseline within 2 years (Barman, Chatterjee and Bhide, 2016). Due to the heterogeneity across different types of ABIs, both in the pathology of injury and loci of damage, the individual may experience a wide variety of possible cognitive impairments (Whyte *et al.*, 2011; Barman, Chatterjee and Bhide, 2016). Cognitive impairments in ABI can have a significant impact on behavioural, social and emotional functioning (Spitz *et al.*, 2012).

The assessment of cognitive impairment may range from brief measures of impairment of consciousness, such as the Glasgow Coma Scale (Teasdale and Jennett, 1974), to more detailed cognitive screening, or a full neuropsychological assessment. A neuropsychological assessment can be time and resource intensive, and is not feasible to complete on every patient (Teager *et al.*, 2020). Routine cognitive screening is a cost-effective and efficient means of identifying which patients may benefit from further in-depth assessment, and which cognitive domains require additional investigation (NICE, 2014; Teager *et al.*, 2020). Results on screening can be used to indicate areas of strengths and needs, which may become targets for future intervention after further assessment (Scottish Acquired Brain Injury Network, 2017). Therefore, cognitive screening tools are essential tools for clinical services assessing patients with ABI.

A number of existing screening tools, including the Mini Mental State Examination (MMSE) (Folstein, Robins and Helzer, 1983) and the Montreal Cognitive Assessment (MoCA)(Nasreddine *et al.*, 2005) have been suggested for use in ABI populations. As shown in the systematic review in Chapter 1, these tools, particularly the MoCA, have the greatest quantity of psychometric evidence (comparatively to other measures) for their use in traumatic

brain injury (TBI) populations specifically. However, evidence for some psychometric properties is either not available at present; or is sparse and requiring replication. Tools such as the MMSE show some promise but suffer from insufficient structural validity and mixed evidence of criterion validity. Additionally the literature as a whole has a number of methodological issues and poor reporting of psychometric properties. Similarly, recent reviews in stroke populations have concluded that there is insufficient, good quality evidence for any cognitive screening tool to be considered a "gold standard" for use (Stolwyk et al., 2014). There is also a clinical demand for measures which can be accessed free of cost, which may impact the utility of some measured identified in recent reviews (Teager *et al.*, 2020).

The Addenbrooke's Cognitive Examination (ACE) is a commonly used cognitive screening measure which provides an index of many of the key areas of brain function (Cullen *et al.*, 2007). The third edition (ACE-III), published by Hsieh et al., (2013), has found widespread use amongst clinicians working with degenerative conditions. Administration takes 15-20 minutes and gives rise to a total ACE-III score as well as separate scores for Attention; Memory; Fluency (often taken to be a crude measure of executive function); Language; and Visuospatial functioning. The ACE-III had initial validation in persons with fronto-temporal dementia and Alzheimer's disease (Hsieh *et al.*, 2013) and has since been validated in alcohol-related brain damage (Brown *et al.*, 2019) and neurocognitive disorder in Parkinson's disease populations (Lucza *et al.*, 2018). The ACE-III has been found to be correlated with deteriorations in functional abilities in those with dementia (Giebel and Challis, 2017). However, its relationship to functional ability in brain injury is unclear (McGhee, Psaila and Allanson, 2017).

The ACE-III is a commonly used measure in brain injury settings because it is quick to administer, free, assesses multiple cognitive domains and has face validity (Teager *et al.*, 2020). It is recommended as a screening tool in ABI by the Scottish Acquired Brain Injury Network (SABIN) (Scottish Aquired Brain Injury Network, 2017). Furthermore, an earlier version of the ACE-III, the ACE-R, has demonstrated superior sensitivity to the MMSE to detect cognitive impairment in brain injury (Gaber, 2008). Given the potential utility of the ACE-III in persons with ABI, and its current use in services, it is important to clarify its validity. It would also be useful to determine whether it is a valid tool to detect disability in this population. Information on the validity of the ACE-III would assist services in making an informed decision about its use in brain injury settings, and may inform the refinement of

clinical guidelines, as many major national guidelines do not recommend any particular cognitive screening tool as part of a cognitive assessment pathway (NICE, 2014). As the ACE-III does not have an "executive functioning" domain, the findings could also be used to determine the suitability of the fluency domain of the ACE-III to predict executive functioning impairment; a domain often impacted in ABI populations. The validity of the fluency domain in this regard does not appear to have been determined by the previous literature (Hsieh *et al.*, 2013).

Aims

The primary aim was to determine the validity of the ACE-III in an ABI population. Performance on each domain of the ACE-III (attention, memory, fluency, language and visuospatial functioning) was compared against standard cognitive tests in the corresponding domain, to determine its concurrent validity. The study also aimed to determine the optimal cut-off for sensitivity and specificity on the ACE-III to detect impairment in the aforementioned cognitive domains. A secondary aim was to explore the relationship between the scores on the domains of the ACE-III and the indexes of the Mayo Portland Adaptability Inventory (MPAI-4), a measure of functional ability, in ABI.

Research Questions

Primary Questions

1. Is the ACE-III a valid screening tool for use in ABI?
    a. Is there a significant correlation of at least moderate magnitude (r >= 0.3) between each domain on the ACE-III and each standard cognitive test in the corresponding domain?
2. Is the ACE-III a sensitive and specific tool (classification performance in each cognitive domain "fair" or better [area under the curve 0.7 or above [Safari et al., 2016]]) to detect cognitive impairment in ABI?

Secondary Questions

3. Is a patient's performance on the ACE-III indicative of their level of functional ability?
    a. What are the magnitudes of correlations between each of the domains (and total score) of the ACE-III and the index scores (and total score) of the MPAI-4?

Method

The full research proposal for this project is available via the link in Appendix 2.0.

Design

This quantitative study used a cross-sectional design in a single clinical centre.

Participants

Participant data were accessed from existing routine clinical data at Graham Anderson House (GAH) in Glasgow, a specialist neuro-behavioural assessment and rehabilitation centre for people with an ABI, operated by the Brain Injury Rehabilitation Trust (BIRT), part of The Disabilities Trust (DT).

Inclusion and Exclusion Criteria

Participants were included in the research database if they had completed the ACE-III upon admission to GAH, and this data was available. The eligible date range was between 2013 (when the ACE-III was published) and the date of data extraction (May 2022). In line with the clinical criteria for admission to GAH, all participants had a diagnosis of moderate-severe ABI, were aged 16-84, and were taking part in inpatient rehabilitation in GAH (types of ABI experienced may include: traumatic brain injury; cerebrovascular accident; hypoxic brain injury; toxic brain injury; infection; or space occupying lesion). Participants were medically stable and did not have a degenerative illness. If participants had been readmitted, their assessment data from their initial admission was used and their readmission data was not included.

Ethics, Governance and Data Processing

Ethical approval was provided from the West of Scotland NHS Research Ethics Committee on 04/10/2021 (Appendix 2.1). The first author's employing board (NHS Ayrshire and Arran) confirmed that Caldicott Guardian approval was not required from the NHS (Appendix 2.2). Two Data Protection Impact Assessments (DPIA) were completed; one for the University of Glasgow and one for The Disabilities Trust (related correspondance in Appendixes 2.3 and 2.4). By signing the DPIA, The Disabilities Trust confirmed that Management approval and Caldicott Guardian approval was provided, as per their current procedures (Appendix 2.5). An information sharing agreement was created and signed by the data protection officer from the Disabilities Trust and the contracts office of the University of Glasgow (Appendix 2.6).

Routine clinical data was accessed retrospectively by the clinical team at GAH. The clinical database was stored on networks at GAH only. Patients consented on admission to their data being used for service evaluation, but not explicitly for the purposes of this research study; this context was considered when planning the governance arrangements for the use of the data for this project, which were deemed by the ethics committee to be appropriate. Patient data, including neuropsychological assessments, the ACE-III, the MPAI-4 and demographic data was collated in a clinical electronic database by GAH staff. For the purposes of this study, additional entries were added to the electronic patient dataset using corresponding information in patient paper files by clinical or admin team at GAH. A research database was then created from this by Disabilities Trust staff with anonymised data. Each participant was given a unique ID number for the research database, which was not traceable back to the patient in the clinical database. No identifiable information was included in the research database or accessed at any point by the trainee or her University supervisor. An ID log was not retained, given the absence of explicit consent. Research and clinical databases were stored separately at GAH. After being securely transferred by Disabilities Trust staff to the trainee, the anonymised research database was stored on a University encrypted laptop and backed up onto secure University networks. All data was handled in line with the Data Protection Act (2018) and General Data Protection Legislation.

Measures

Demographic and Clinical Variables

Demographic and clinical variables recorded included gender, age at admission, age at time of injury, number of days between injury and admission and the patient's diagnosis. It was recorded if the patient had (yes/no) experienced pre-injury psychosis, drug-dependence, alcohol abuse, multiple traumas (any other injury apart from the brain injury, which might have been sustained as a result of a road traffic accident, fall, attack etc., including; broken bones, spinal cord, injuries to other organs such as the lungs or spleen) or has another medical condition. Data on Scottish Index of Multiple Deprivation (SIMD) quintile (Scottish Government, 2020) was sought, but was not available. Each patient's Hospital Anxiety and Depression Scale (HADS) score was included (a valid measure of anxiety and depression in brain injury populations, with good sensitivity and specificity for both (>0.8); Dahm et al., 2013).

Addenbrooke's Cognitive Examination (ACE-III)

The ACE-III is a screening tool for cognitive functioning, which has been validated for use in dementia patients (Hsieh *et al.*, 2013). It gives a total score (/100), as well as subscale scores for Attention (/18); Memory (/26); Fluency (/14); Language (/26); and Visuospatial functioning (/16). At GAH, the ACE-III is typically administered during the pre-admission process (2-3 weeks prior to admission), typically by a Clinical Psychologist or other trained professional. In some cases, it is administered shortly after admission.

Mayo Portland Adaptability Index (MPAI-4)

The MPAI-4 is a measure of functional ability, validated in brain injury samples (Bellon, Malec and Kolakowsky-Hayner, 2012; Malec *et al.*, 2012). It consists of 35 items, each ranging from 0 to 4. The first 29 items contribute to three subscales: the ability index (the degree of impairment experienced); the adjustment index (how the patient is adjusting to changes in function); and the participation index (ability of the individual to function and carry out everyday tasks). The remaining 6 items detail pre- and post-injury information and are not used

in the final scoring. The MPAI-4 is routinely completed at initial assessment on entry to GAH. This measure was completed by the clinician.

Neuropsychological Assessment Battery

At admission to GAH, a neuropsychological assessment battery is administered to patients by a Clinical Psychologist or other trained professional. It contains a variety of possible cognitive tests and those administered vary between patients (Table 2.1). The tests have been mapped by the present research team to the primary ACE-III domain scores of interest, although it is acknowledged that there is overlap between domains e.g. attention and executive function. Some patients may have also been administered the BIRT Memory and Information Processing Battery [BMIPB] but data was not available (Coughlan, Oddy and Crawford, 2007).

Table 2. 1 Cognitive Tests

Measure	Examples of Validation Studies in ABI	Sub-scales/ Subtests	Domain Assessed	Corresponding ACE-III Domain
Test of Everyday Attention (TEA)(Robertson *et al.*, 1994) –	Evidence for it's structural and discriminative validity (Robertson *et al.*, 1996; Bate, Mathias and Crawford, 2001)	Map Search	Visual Selective Attention	Attention
		Elevator Counting	Sustained attention	
		Elevator Counting (with distraction)	Sustained attention and auditory verbal working memory	
		Visual Elevator	Attentional switching	
		Auditory Elevator (with Reversal)	Attentional switching	
		Telephone Search	Visual selective attention	
		Telephone Search (Dual Task)	Sustained attention and divided attention	
		Lottery	Sustained attention	
Galveston Orientation and Amnesia Test	Evidence of convergent and predictive	7 items (orientation); 3 items allow	Temporal orientation and PTA/RA	Attention (orientation)

Table 2. 1 Cognitive Tests

(GOAT)(Levin, O'donnell and Grossman, 1979)	validity (Levin, O'donnell and Grossman, 1979)	calculation of post-traumatic amnesia (PTA)/retrograde amnesia (RA)		
Rivermead Behavioural Memory Test (RBMT)(Wilson *et al.*, 2008)	Evidence of internal consistency, structural and construct validity (Küçükdeveci *et al.*, 2008)	Items 1 & 2 (name)	Memory/orientation	Memory
		Items 3, 4, 8, 9, 10	Behavioural memory	
		Items 5, 6, 7	Visual memory	
		Item 11	orientation	
		Item 12	Date	
Behavioural Assessment of Dysexecutive Syndrome (BADS)(Wilson *et al.*, 1996)	Evidence of convergent validity with clinician ratings and discriminant validity (Bennett, Ong and Ponsford, 2005; Boelen *et al.*, 2009)	Temporal Judgement	Executive functioning	Fluency
		Rule Shift		
		Action Program		
		Key Search		
		Zoo Map		
		Modified Six Elements		
		Dysexecutive Questionnaire		
Wechsler Adult Intelligence Scale (WAIS-IV) (Wechsler, 2008)	Evidence for discriminant validity and criterion validity (Theiling, Petermann and Daseking, 2013; Carlozzi *et al.*, 2015; Donders and Strong, 2015)	Verbal Comprehension Index (VCI)	Language	Language
		Working Memory Index (WMI)	Working Memory	Attention
		Perceptual Reasoning Index (PRI)	Visuospatial	Visuospatial

A summary of cognitive tests and their sub-scales/components; the cognitive domains that they relate to; and the primary corresponding domain on the ACE-III

Data Analysis

All data were analysed using Microsoft Excel and IBM SPSS statistics version 28. Descriptive statistics, including for the aforementioned demographic and clinical variables, are reported. As research ethics approval was only given for data to be accessed for those with ACE-III data, the research team was unable to directly examine data comparisons between patients with and without ACE-III data; however, summary information on the characteristics of the whole GAH population are noted in the results section below for context. Patterns of missingness on the variables within the sample were explored. The data were inspected for possible errors (e.g. out-of-range scores) and any such erroneous scores were deleted.

Patients' test scores from the neuropsychological assessment battery were converted to z-scores, where possible, using published normative data. Impairment was defined as a score of $z<-1.64$ (equivalent to 5^{th} percentile), using criteria in Strauss et al., (2006). The patient's z-score for each cognitive test was coded in the dataset as impaired/not impaired.

To determine validity, the five sub-scores on the ACE-III were correlated separately with the z-scores of each corresponding cognitive test, where available (Table 2.1). Normality tests were conducted, and a visual exploration of the data revealed that assumptions of parametric analysis were not met. Spearman's correlations were therefore used for all comparisons. All spearman correlations were interpreted using proposed criteria in Prion and Haerling, (2014). Attention (ACE-III) sub-scale score was correlated with GOAT (total score), TEA scores and WAIS-IV WMI. Memory (ACE-III) sub-scale score was correlated with the RBMT (total score). Fluency (ACE-III) sub-scale score was correlated with the BADS (total score and subtests). Language (ACE-III) sub-scale score was correlated with the WAIS-IV VCI. The Visuospatial (ACE-III) sub-scale score was correlated with the WAIS-IV PRI. For cases where a patient had completed two different cognitive tests within the same domain (e.g., TEA and GOAT, for ACE-III attention subtest), they were included in both correlation analyses. The ACE-III Total score was also correlated with the WAIS-IV FSIQ.

A sensitivity analysis was conducted for the above correlation analyses, in which the ACE-III was corrected for age using a regression method. This was to account for the fact that the cognitive test z-scores are age-corrected while the raw ACE-III scores are not. There was no difference in the correlations which met statistical significance between non age-corrected and

age-corrected scores. A small number of correlations changed magnitude (6/19 ACE x cognitive test correlations changed by r>0.1; and 2/19 changed by r>0.2). Therefore, non-age corrected values are reported.

Receiver operating characteristic (ROC) curves were utilised to visualise the ability of ACE-III sub-scores to detect impairment status (impaired vs not impaired) in each cognitive domain. AUC's (areas under the curve) were calculated; and sensitivity and specificity reported for adequately powered tests.

The total ACE-III score, along with the five subscales, were correlated against the total MPAI-4 score, along with the three indices. This resulted in 24 correlations. These were not corrected for multiple comparisons, as this was a secondary research question and the focus was primarily on the size of the correlations rather than statistical significance.

Analyses for the primary research questions were not corrected for multiple comparisons, as each correlation utilises different test score data to examine validity in different cognitive domains. An exception was made where two or more correlations were conducted within a particular domain (e.g., ACE-III Attention score vs GOAT, and ACE-III Attention score vs TEA), in which case the significance levels were adjusted accordingly (e.g., $p<0.025$ if two tests are conducted in one domain).

Sample size and Power Calculation

Initial information from GAH indicated there was an estimated potential sample size of n=100-280. It was calculated that for correlation analyses, this sample size would give power to detect a small to medium correlation: G*Power indicated that a sample of n=100 would enable reliable detection of a correlation of r=0.28 or larger with 80% power and $p<0.05$ (two-tailed) (Faul *et al.*, 2009). For the ROC curve analysis, MedCalc (https://www.medcalc.org/) indicated that the estimated minimum sample size (n=100) would allow for the detection of an area under the curve (AUC) of 0.7 (assuming the ratio of sample sizes of each group is <3.75). The maximum sample size could allow for the detection of an AUC of 0.60 (assuming the ratio of each group is <1.5).

The actual sample size in the final dataset was lower than expected (n=26), with varying degrees of missingness within each variable. Power analyses ran in G*Power indicated correlation sample size for neuropsychological tests (n=5-21) had the power to detect between a medium -large correlation (critical r=0.43-0.88). MPAI x ACE-III correlations had the power to detect a medium correlation (critical r= 0.40-0.42). For the ROC curve analysis: MedCalc indicated that available sample size allowed for sufficient power (0.80) to detect an AUC of between 0.85 to 0.96 (depending on the comparison). The ACE-III x RBMT sample size did not allow for the detection of the maximum possible AUC of 1.0 and so this analysis was not conducted.

Results

Sample characteristics

A total of 26 participants were included in analyses. Demographic and clinical characteristics for participants are reported in Table 2.2. The sample was predominantly male (80.77%) with a a mean age of 46.5 years old (SD=13.73) on admission to GAH. The most common cause of ABI was TBI (57.7%); and the majority of the sample had a history of pre-morbid alcohol abuse (53.8%). Data held centrally at GAH for all patients indicated that the study sample was similar to the wider patient population with regard to age (M=47, SD=14), HADS Anxiety (M=8, SD=5), HADS Depression (M=6, SD=5), but different with regard to MPAI (raw score) (Total M=72, SD=14) and days since injury (M=1011 days [or 2.76 years]; SD=2405).

Table 2. 2 Sample Demographic and Clinical Characteristics

	Data available (n)		
Sex *n (%)*	26	Males Female	21 (80.77%) 5 (19.23%)
Age in years (on admission) *Mean (SD)*	26	46.5 (13.73)	
Age in years (at injury) *Mean (SD)*	25	43.08 (16.34)	

Table 2. 2 Sample Demographic and Clinical Characteristics

	Data available (n)		
Days between injury and admission *Mean (SD)*	25	1143.40 (2222.70) or 3.13 years	
Diagnosis Type *n (%)*	26	TBI CVA Hypoxia Infection Other	15 (57.7%) 5 (19.2%) 2 (7.7%) 2 (7.7%) 2 (7.7%)
Comorbidities *n (%)*	25	Pre-injury Psychosis Pre-injury Drug Dependence Pre-injury Alcohol Abuse Multiple Trauma Other Medical	0 4 (15.4%) 14 (53.8%) 6 (23.1%) 9 (34.6%)
HADS *Mean (SD)*	24	HADS A HADS D	7.54 (5.23) 6.63 (4.04)
MPAI-4 (T-score) *Mean (SD)*	25	Abilities Adjustment Participation Total	50.48 (7.33) 54.68 (4.39) 54.64 (7.93) 57.96 (6.96)

Abbreviations: CVA, cerebrovascular accident; HADS A, Hospital Anxiety and Depression Scale for Anxiety; HADS D, HADS for Depression; MPAI-4, Mayo Portland Adaptability Index-4 (reported in t-score format); SD, standard deviation; TBI, traumatic brain injury;

ACE-III and Neuropsychological Tests

Descriptive and missingness statistics for the ACE-III and each neuropsychological variable are reported in Table 2.3. One participant had a missing ACE-III total score, but had several subscale scores, and so was included in relevant analyses. Missingness in the neuropsychological test data ranged from 15.4% (WAIS-PRI) to 76.9% (TEA lottery) depending on the variable.

Table 2. 3 ACE-III and Neuropsychological Assessment Data

		N (% of full sample of 26)	Median (IQR)	Impaired n (% of those with data)
ACE-III	Attention	24 (92.3%)	13.50 (7.00)	
	Memory	24 (92.3%)	14.50 (7.00)	

	Fluency	24 (92.3%)	7.00 (5.50)	
	Language	23 (88.5%)	17.00 (15.00)	
	Visuospatial	23 (88.5%)	13.00 (7.00)	
	TOTAL	25 (96.2%)	64.00 (32.50)	
TEA	Maps (1 minute) (SS)	10 (38.46%)	3.00 (5.00)	9 (90.00%)
	Maps (2 minute) (SS)	10 (38.46%)	2.00 (4.00)	7 (70.00%)
	Elevator (counting)	13 (50.00%)	7.00 (1.00)	7 (53.85%)
	Elevator (distraction) (SS)	12 (46.15%)	8.00 (5.50)	4 (33.33%)
	Lottery (SS)	6 (23.08%)	8.00 (5.00)	2 (33.33%)
GOAT	TOTAL	12 (46.15%)	88.00 (29.00)	4 (33.33%)
RBMT	GMI (standard score)	12 (46.15%)	60.50 (9.00)	11 (92.67%)
BADS	Rule Shift	18 (69.20%)	2.50 (3.25)	
	Action Program	19 (73.10%)	2.00 (3.00)	
	Key Search	19 (73.10%)	2.00 (2.00)	
	Temporal Judgement	18 (69.20%)	2.00 (1.25)	
	Zoo Map	17 (65.40%)	1.00 (1.50)	
	Modified Six Elements	15 (57.7%)	2.00 (1.00)	
	Full Scale (standard score)	16 (61.54%)	68.00 (31.00)	11 (68.75%)
WAIS (standard scores)	VCI	21 (80.77%)	76.00 (14.00)	10 (47.62%)
	PRI	22 (84.62%)	81.00 (19.00)	7 (31.82%)
	WMI	21 (80.77%)	80.00 (15.00)	8 (38.10%)
	PSI	21 (80.77%)	68.00 (17.00)	15 (71.43%)
	FSIQ	21 (80.77%)	74.00 (12.00)	14 (66.67%)

Abbreviations. ACE-III, Addenbrookes Cognitive Exam-III; BADS Behavioural Assessment of Dysexecutive Syndrome ; GOAT Galveston Orientation and Amnesia Test; RBMT Rivermead Behavioural Memory Test; TEA Test of Everyday Attention (TEA); WAIS, Wechsler Adult Intelligence Scale (FSIQ, full scale intelligence quotient; PRI, perceptual reasoning index; PSI processing speed index; VCI, verbal comprehension index; WMI, working memory index).

n.b test scores are in raw score format, unless otherwise specified (e.g. as scaled score [SS] or standard score); impaired for the TEA elevator counting was defined as <7 (TEA manual states that a score of 6 = "doubtful" and scores of 5 = "definitely abnormal")

A summary of all Spearman correlation analyses is reported in Tables 2.4 -2.9. Scatterplots of correlations can be found in Appendix 2.7. There was a strong correlation between the ACE-III Attention subtest and WAIS-IV WMI, which had a wide confidence interval. No significant correlations were found between the ACE-III attention domain and the TEA subtests or GOAT score (Table 4). The ACE-III Memory X RBMT correlation did not meet significance (Table 5.). No ACE-III Fluency domain X BADS correlations met the significance threshold (Table 6.).

Table 2. 4 Correlations between ACE-III Attention and Neuropsychological Test Score

	n	ρ	p	CI (95%)
TEA Maps 1	9	-0.01	0.98	-0.68 to 0.67
TEA Maps 2	9	-0.57	0.11	-0.90 to 0.18
TEA Elevator Distraction	11	0.39	0.23	-0.29 to 0.81
TEA Lottery	5	0.89	0.04	0.02 to 0.99
GOAT (Raw Score)	11	0.67	0.02	0.10 to 0.91
WAIS-IV WMI	19	0.60*	0.007	0.18 to 0.83

Tests: ACE-III, Addenbrookes Cognitive Exam-III; GOAT Galveston Orientation and Amnesia Test; TEA Test of Everyday Attention (TEA); WAIS-IV, Wechsler Adult Intelligence Scale Version 4 (WMI, working memory index).

*ρ, spearman's Rho; CI, confidence interval. * test meets significance threshold (p<0.008, corrected for multiple comparisons)*
Correlations are reported for the ACE-III raw score X each neuropsychological test (z scores; unless specificed i.e. GOAT scores were correlated as a raw score)

Table 2. 5 Correlations between ACE-III Memory and Neuropsychological Test

	n	ρ	p	CI (95%)
RBMT Total	11	-0.05	0.883	-0.64 to 0.58

ACE-III, Addenbrookes Cognitive Exam-III; RBMT=Rivermead Behavioural Memory Test
*ρ, spearman's Rho; CI, confidence interval. * = test meets significance threshold (p<0.05).*

Correlations are reported for the ACE-III raw score X each neuropsychological test (z scores).

Table 2. 6 Correlations between ACE-III Fluency and Neuropsychological Tests

	n	ρ	p	CI (95%)
BADS Rule Shift (Raw Score)	17	0.43	0.09	-0.79 to 0.76
BADS Action Program (Raw Score)	17	0.39	0.13	-0.13 to 0.74
BADS Key Search (Raw Score)	17	0.48	0.05	-0.02 to 0.78
BADS Temporal Judgement (Raw Score)	17	0.30	0.24	-0.22 to 0.69
BADS Zoo Map (Raw Score)	16	0.52	0.04	0.02 to 0.81
BADS Modified Six Elements (Raw Score)	14	0.01	0.98	-0.54 to 0.55
BADS Full Scale	15	0.54	0.04	0.03 to 0.83

ACE-III, Addenbrookes Cognitive Exam-III; BADS Behavioural Assessment of Dysexecutive Syndrome

ρ, spearman's Rho; CI, confidence interval. No tests met significance threshold (p<0.007, adjusted for multiple comparisons).

Correlations are reported for the ACE-III raw score X each neuropsychological test (z scores).

A moderate correlation was found between ACE-III Language subtest scores and the WAIS VCI (Table 7). There was a strong correlation found between ACE-III Visuospatial subtest and the WAIS PRI (Table 8). There was a moderate correlation between ACE-III total score and WAIS FSIQ (Table 9). However, all correlations had wide confidence intervals.

Table 2. 7 Correlation between ACE-III Language and Neuropsychological Test

	n	ρ	p	CI (95%)
WAIS VCI	18	0.51*	0.03	0.04 to 0.79

ACE-III, Addenbrookes Cognitive Exam-III; WAIS VCI= Weschler Adult Intelligence Scale Verbal Comprehension Index
*ρ, spearman's Rho; CI, confidence interval. * = test meets significance threshold (p<0.05).*
Correlations are reported for the ACE-III raw score X each neuropsychological test (z scores).

Table 2. 8 Correlation between ACE-III Visuospatial and Neuropsychological Test

	n	ρ	p	CI (95%)
WAIS PRI	19	0.62*	<0.01	0.21 to 0.84

ACE-III, Addenbrookes Cognitive Exam-III; WAIS PRI = Weschler Adult Intelligence Scale Perceptual Reasoning Index * *ρ, spearman's Rho; CI, confidence interval. * = test meets significance threshold (p<0.05).*

Correlations are reported for the ACE-III raw score X each neuropsychological test (z scores).

Table 2. 9 Correlation between ACE-III Total Score and WAIS-IV FSIQ

	n	ρ	p	CI (95%)
WAIS-IV FSIQ	20	0.58*	<0.01	0.17-0.82

*ACE-III, Addenbrookes Cognitive Exam-III; WAIS, Wechsler Adult Intelligence Scale (FSIQ, full scale intelligence quotient. ρ, spearman's Rho; CI, confidence interval. * = test meets significance threshold (p<0.05).*
Correlations are reported for the ACE-III raw score X each neuropsychological test (z scores).

The ability of the ACE-III score to detect whether or not participants were in the impaired range on the reference test was assessed using ROC curve analysis (Table 2.10). The minimum AUC which the present sample size could be expected to reliably detect was 0.85. Therefore, the only comparison which was adequately powered and met significance was ACE-III Visuospatial x WAIS-PRI. Participants performance on the ACE-III Visuospatial subtest provided the ability to classify those who were impaired/unimpaired on the WAIS-PRI, with a "very good" AUC. A cut-off of >11/16 on the visuospatial subtest resulted in an acceptable balance of sensitivity (75.0%) and specificity (85.7%) to detect impairment on the WAIS-PRI (Power, Fell and Wright, 2013). ROC curves can be found in Appendix 2.8.

.

Table 2. 10 ROC Curve Analysis

ACE-III Subtest	Neuropsychological Test	n	AUC	p	Standard Error	CI (95%)
ACE-III Attention	GOAT	11	0.83	0.03	0.16	0.53-1.14
	WAIS-IV WMI	19	0.78	0.01	0.12	0.56-1.01
ACE-III Fluency	BADS (Full Scale)	15	0.79	0.03	0.13	0.53 to 1.05
ACE-III Language	WAIS-IV VCI	18	0.76	0.03	0.12	0.53 to 0.99
ACE-III Visuospatial	WAIS-IV PRI	19	0.89	<0.001	0.76	0.74 to 1.04

Table 2. 10 ROC Curve Analysis

ACE-III Subtest	Neuropsychological Test	n	AUC	p	Standard Error	CI (95%)
ACE-III Total	WAIS-IV FSIQ[a]	20	0.69	-	-	-

ACE-III, Addenbrookes Cognitive Exam-III; BADS Behavioural Assessment of Dysexecutive Syndrome ; GOAT Galveston Orientation and Amnesia Test; RBMT Rivermead Behavioural Memory Test; TEA Test of Everyday Attention (TEA); WAIS, Wechsler Adult Intelligence Scale (FSIQ, full scale intelligence quotient; PRI, perceptual reasoning index; PSI processing speed index; VCI, verbal comprehension index; WMI, working memory index).

AUC = area under the curve; CI, confidence interval; ROC, Receiver Operating Curve Analysis.

a = full ROC analyses not possible "at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased."

ACE-III and MPAI-4

To address the secondary research question, patients' MPAI-4 scores (where lower scores indicate better functioning) were correlated with their ACE-III scores (where higher scores indicate better functioning (Table 2.11) Of note, the ACE-III Visuospatial domain was strongly, negatively correlated with MPAI-4 ability and MPAI-4 total scores. The correlations in other domains were typically negligible, weak or moderate in strength; and negative in direction. Many comparisons did not meet significance, and there were wide confidence intervals across all comparisons.

Table 2. 11 Correlations (ρ and 95% CI) between ACE-III and MPAI-4.

	ACE-III Attention	ACE-III Memory	ACE-III Fluency	ACE-III Language	ACE-III Visuospatial	ACE_III total
MPAI-4 Ability	-0.37 (-0.69 to 0.06), p=0.08	-0.19; (-0.57 to 0.25) p=0.39	-0.29; (-0.63 to 0.15) p=0.18	-0.34; (-0.67 to 0.11) p=0.12	-0.63*; (-0.84 to -0.28) p<0.01	-0.39; (-0.69 to 0.03) p=0.06
MPAI-Adjustment	-0.24 (-0.60 to 0.20), p=0.27	0.07; (-0.37 to 0.48) p=0.77	-0.31; (-0.65 to 0.13) p=0.15	-0.08; (-0.37 to 0.49) p=0.74	-0.29; (-0.64 to 0.16) p=0.19	-0.16; (-0.54 to 0.28) p=0.47
MPAI-4 Participation	-0.46* (-0.74 to -0.04), p=0.03	-0.77; (-0.48 to 0.36) p=0.73	-0.50*; (-0.76 to -0.10) p=0.02	-0.22; (-0.60 to 0.23) p=0.32	-0.53*; (-0.76 to -0.13) p=0.01	-0.41*; (-0.70 to 0.01) p=0.048
MPAI-4 Total	-0.46*; (-0.74 to -0.05) p=0.03	-0.166; (-0.55 to 0.28) p=0.45	-0.43*; (-0.73 to -0.02) p=0.03	-0.32; (-0.66 to 0.13) p=0.15	-0.68*; (-0.86 to -0.35) p<0.001	-0.45*; (-0.73 to -0.04) p=0.03

*ACE-III, Addenbrookes Cognitive Exam-III; MPAI-4, Mayo Portland Adaptability Index-4. ρ, spearman's Rho; CI, confidence interval. * = test meets significance threshold (p<0.05).*

Discussion

Given data availability and missingness within the dataset, this project was estimated to have sufficient power to detect only an excellent-outstanding AUC; or a moderate-very strong correlation (depending on sample size of the comparison). Nevertheless, this study was able to make a valuable contribution to the evidence base by providing preliminary support for the ACE-III's convergent validity with standard cognitive tests and functional measures; as well as it's criterion validity to predict cognitive impairment. For patients with TBI, their ACE-III subtest scores were moderately-strongly correlated with their corresponding WAIS-IV index scores, for the attention, language and visuospatial domains. This suggests that a stronger performance on these domains on the ACE-III may be associated with stronger performance

on standard neuropsychological tests. Several ACE-III subtest scores (attention, fluency and visuospatial) and the ACE-III total score were moderately-strongly correlated with the MPAI, suggesting worse performance on several domains of the ACE-III may be associated with functional impairment. However, wide confidence intervals suggests these findings are imprecise. Finally patient's scores on the ACE-III visuospatial subtest were able to classify patients with/without cognitive impairment in the visuospatial domain. While these findings should be considered within the context of the limitations (discussed below), they provide preliminary support for the validity ACE-III as a cognitive screening tool in TBI, particularly to detect visuospatial impairment. This has implications for the ACE-III's use in post-acute, rehabilitation settings. A patient scoring poorly on certain ACE-III subtests are likely to score poorly on standard tests for the corresponding domain (for attention, language and visuospatial domains). Patients scoring poorly in the visuospatial domain should be provided additional assessment and intervention for possible visuospatial impairment.

The findings of this study should be considered in the context of several key limitations. First, while several moderate-strong correlations were found, the confidence intervals were typically large, suggesting the correlations reported are an imprecise estimate of the true correlation sizes. Additionally, while many ROC curve analyses appeared promising (with AUC's >0.7), many require additional research in with a larger sample size to clarify whether ACE-III subtest scores can classify cognitive impairment (in attention, memory, fluency and language domains). For comparisons within attention, memory and fluency domains, there was insufficient power in the current study to detect weak or moderate correlations (depending on the sample size per comparison). The ACE-III was originally designed and validated as a measure for use in dementia and has less of a focus on executive function (only one subtest on fluency). In contrast, executive function is a key domain, commonly impacted in ABI (Hsieh *et al.*, 2013; Barman, Chatterjee and Bhide, 2016). As such, some have doubted its face validity as cognitive screen in ABI populations (Whyte *et al.*, 2011; Barman, Chatterjee and Bhide, 2016). Therefore, it is essential that adequately powered studies further clarify the validity of the ACE-III fluency domain as an indicator of executive function, before the ACE-III can be recommended as a cognitive screen. Divergent validity was not assessed as part of this study, for example, by correlating each ACE-III subscale with a neuropsychological test in a cognitive domain which does not correspond directly to the ACE-III subscale domain (such as ACE-III language with an executive function measure). Therefore it is not possible to say, for example, whether the ACE-III visuospatial domain is a specific indicator of visuospatial impairment

only, or rather if it functions as an indicator of impairment across multiple domains. In order to conceptualise the ACE-III as a valid multi-dimensional screening tool, which is able to detect cognitive impairment in discrete domains, it's divergent validity must be clarified.

As discussed, a limitation of this study is that the small sample size and missingness within the data. However, it would not have been appropriate to impute missing values in such a small sample. It is also likely that missing values were not missing at random, as patients may have been unable to complete several tests owing to their low cognitive function. Missingness and errors are typical for routinely collected data, and effort was made to remove values which were clearly inaccurate. However, due to the data protection procedures agreed, it would not have been possible to establish whether there were additional errors within the dataset. The use of secondary data did not allow for the collection of individual test items to allow for calculation of internal consistency; or the creation of a study design that would allow assessment of reliability. The levels of co-morbid alcohol abuse with the sample may appear high, and which may increase the likelihood of those in the sample also possibly experiencing alcohol-related cognitive impairment. However, the proportion of those with an alcohol abuse history appears to be consistent with the demographics of the broader ABI population and it would not have been appropriate to exclude those individuals to maintain the generalisability of the findings (Weil, Corrigan and Karelina, 2018). The mean HADS anxiety score was above clinical threshold level, but again, this is consistent with previous literature in ABI samples (Longworth *et al.*, 2018).

A strength of this study is its accordance with the COSMIN guidelines for statistical reporting and risk of bias in psychometric research (Mokkink *et al.*, 2010). For example, a clear statement was made in the research questions about the relationship between the ACE-III and measures expected to be convergent, psychometric properties of convergent measures were highlighted in the methods section, and appropriate statistics were reported. The patient population also appeared to be broadly similar to the wider post-acute rehabilation population within GAH, indicating it's generalisability.

Before the ACE-III can be recommended as a cognitive screening tool, research with a larger sample size is required to confirm the convergent and criterion validity of this measure in a ABI sample. This would allow increased precision in establishing the true correlation between the ACE-III and standard cognitive tests. Future research designs should also aim to evaluate

the internal consistency, reliability and other forms of validity of the ACE-III in ABI populations, in a manner consistent with COSMIN guidelines (Mokkink *et al.*, 2010). As some of the correlations in this study changed magnitude after age correction, there may be a small amount of age confounding in the results. Therefore, age should be taken into account in sensitivity analyses when conducting future research with the ACE-III in ABI samples. Finally, this small sample was taken from a moderate-severe severity, post-acute rehabilitation setting. Additional research should establish the ACE-III's validity in mild TBI samples and acute settings. This will inform the development of guidelines use of the ACE-II in major trauma centres (Teager *et al.*, 2020).

Conclusions

Despite its wide use in ABI settings in the UK, the ACE-III had yet to be validated in this population. This study is the first to provide preliminary evidence that the ACE-III may be valid for use within this population. Additional psychometric evaluation on the ACE-III, with a focus on an increased sample size and assessment of its reliability, will be necessary to confirm its utility in this clinical population.

References

Barman, A., Chatterjee, A. and Bhide, R. (2016) 'Cognitive Impairment and Rehabilitation Strategies After Traumatic Brain Injury', *Indian Journal of Psychological Medicine*, 38(3), pp. 172–181. Available at: https://doi.org/10.4103/0253-7176.183086.

Bate, A.J., Mathias, J.L. and Crawford, J.R. (2001) 'Performance on the Test of Everyday Attention and Standard Tests of Attention following Severe Traumatic Brain Injury', *The Clinical Neuropsychologist*, 15(3), pp. 405–422. Available at: https://doi.org/10.1076/clin.15.3.405.10279.

Bellon, K., Malec, J.F. and Kolakowsky-Hayner, S.A. (2012) 'Mayo-Portland Adaptability Inventory-4', *The Journal of Head Trauma Rehabilitation*, 27(4), pp. 314–316. Available at: https://doi.org/10.1097/HTR.0b013e3182562f04.

Bennett, P.C., Ong, B. and Ponsford, J. (2005) 'Assessment of executive dysfunction following traumatic brain injury: Comparison of the BADS with other clinical neuropsychological measures', *Journal of the International Neuropsychological Society*, 11(05). Available at: https://doi.org/10.1017/S1355617705050721.

Boelen, D.H.E. *et al.* (2009) 'Executive dysfunction in chronic brain-injured patients: Assessment in outpatient rehabilitation', *Neuropsychological Rehabilitation*, 19(5), pp. 625–644. Available at: https://doi.org/10.1080/09602010802613853.

Brown, P. *et al.* (2019) 'Applicability of the ACE-III and RBANS Cognitive Tests for the Detection of Alcohol-Related Brain Damage', *Frontiers in Psychology*, 10, p. 2636. Available at: https://doi.org/10.3389/fpsyg.2019.02636.

Carlozzi, N.E. *et al.* (2015) 'An Examination of the Wechsler Adult Intelligence Scales, Fourth Edition (WAIS-IV) in Individuals with Complicated Mild, Moderate and Severe Traumatic Brain Injury (TBI)', *The Clinical Neuropsychologist*, 29(1), pp. 21–37. Available at: https://doi.org/10.1080/13854046.2015.1005677.

Coughlan, A.K., Oddy, M. and Crawford, J.R. (2007) *BIRT memory and information processing battery (BMIPB)*. Horsham, UK: Brain Injury Rehabilitation Trust.

Cullen, B. *et al.* (2007) 'A review of screening tests for cognitive impairment', *Journal of Neurology, Neurosurgery & Psychiatry*, 78(8), pp. 790–799. Available at: https://doi.org/10.1136/jnnp.2006.095414.

Dahm, J., Wong, D. and Ponsford, J. (2013) 'Validity of the Depression Anxiety Stress Scales in assessing depression and anxiety following traumatic brain injury', *Journal of Affective Disorders*, 151(1), pp. 392–396. Available at: https://doi.org/10.1016/j.jad.2013.06.011.

Donders, J. and Strong, C.-A.H. (2015) 'Clinical utility of the Wechsler Adult Intelligence Scale-Fourth Edition after traumatic brain injury', *Assessment*, 22(1), pp. 17–22. Available at: https://doi.org/10.1177/1073191114530776.

Faul, F. *et al.* (2009) 'Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses', *Behavior Research Methods*, 41(4), pp. 1149–1160. Available at: https://doi.org/10.3758/BRM.41.4.1149.

Folstein, M.F., Robins, L.N. and Helzer, J.E. (1983) 'The Mini-Mental State Examination', *Archives of General Psychiatry*, 40(7), pp. 812–812. Available at: https://doi.org/10.1001/archpsyc.1983.01790060110016.

Gaber, T.A.-Z.K. (2008) 'Evaluation of the Addenbrooke's Cognitive Examination's validity in a brain injury rehabilitation setting', *Brain Injury*, 22(7–8), pp. 589–593. Available at: https://doi.org/10.1080/02699050802132461.

Giebel, C.M. and Challis, D. (2017) 'Sensitivity of the Mini-Mental State Examination, Montreal Cognitive Assessment and the Addenbrooke's Cognitive Examination III to everyday activity impairments in dementia: an exploratory study', *International Journal of Geriatric Psychiatry*, 32(10), pp. 1085–1093. Available at: https://doi.org/10.1002/gps.4570.

Headway (2017) *ABI 2016-2017 Statistics based on UK admissions*. Available at: https://www.headway.org.uk/about-brain-injury/further-information/statistics/ (Accessed: 2 May 2022).

Hsieh, S. *et al.* (2013) 'Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal Dementia and Alzheimer's Disease', *Dementia and Geriatric Cognitive Disorders*, 36(3–4), pp. 242–250. Available at: https://doi.org/10.1159/000351671.

Küçükdeveci, A.A. *et al.* (2008) 'Construct validity and reliability of the rivermead behavioural memory test in the Turkish population', *Brain Injury*, 22(1), pp. 75–82. Available at: https://doi.org/10.1080/02699050701809011.

Levin, H.S.P.D., O'donnell, V.M.M.A. and Grossman, R.G.M.D. (1979) 'The Galveston Orientation and Amnesia Test: A Practical Scale to Assess Cognition after Head Injury', *Journal of Nervous*, 167(11), pp. 675–684.

Longworth, C. *et al.* (2018) 'The nature of self-esteem and its relationship to anxiety and depression in adult acquired brain injury', *Neuropsychological Rehabilitation*, 28(7), pp. 1078–1094. Available at: https://doi.org/10.1080/09602011.2016.1226185.

Lucza, T. *et al.* (2018) 'Comparing Sensitivity and Specificity of Addenbrooke's Cognitive Examination-I, III and Mini-Addenbrooke's Cognitive Examination in Parkinson's Disease', *Behavioural Neurology*, p. 10.

Malec, J.F. *et al.* (2012) 'Mayo-Portland Adaptability Inventory: Comparing Psychometrics in Cerebrovascular Accident to Traumatic Brain Injury', *Archives of Physical Medicine and Rehabilitation*, 93(12), pp. 2271–2275. Available at: https://doi.org/10.1016/j.apmr.2012.06.013.

McGhee, W., Psaila, K. and Allanson, J. (2017) 'Evaluation of the Addenbrooke's Cognitive Examination-III and its relationship with functional ability after traumatic brain injury', *Clinical Rehabilitation*, 31(12), pp. 1688–1688. Available at: https://doi.org/10.1177/0269215517734025.

Mokkink, L.B. *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), p. 22. Available at: https://doi.org/10.1186/1471-2288-10-22.

Nasreddine, Z.S. *et al.* (2005) 'The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment', *Journal of the American Geriatrics Society*, 53(4), pp. 695–699. Available at: https://doi.org/10.1111/j.1532-5415.2005.53221.x.

NICE (2014) *Clinical guideline [CG176]. Research recommendations. Head injury: assessment and early management.* NICE. Available at: https://www.nice.org.uk/guidance/cg176/chapter/2-Research-recommendations (Accessed: 13 December 2020).

Power, M., Fell, G. and Wright, M. (2013) 'Principles for high-quality, high-value testing', *BMJ Evidence-Based Medicine*, 18(1), pp. 5–10. Available at: https://doi.org/10.1136/eb-2012-100645.

Prion, S. and Haerling, K.A. (2014) 'Making Sense of Methods and Measurement: Pearson Product-Moment Correlation Coefficient', *Clinical Simulation In Nursing*, 10(11), pp. 587–588. Available at: https://doi.org/10.1016/j.ecns.2014.07.010.

Robertson, I. *et al.* (1994) 'The Test of Everyday Attention. Manual', *Test Reviews*, 4, pp. 51–55.

Robertson, I.H. *et al.* (1996) 'The structure of normal human attention: The Test of Everyday Attention', *Journal of the International Neuropsychological Society*, 2(6), pp. 525–534. Available at: https://doi.org/10.1017/S1355617700001697.

Safari, S. *et al.* (2016) 'Evidence Based Emergency Medicine; Part 5 Receiver Operating Curve and Area under the Curve', *Emergency*, 4(2), pp. 111–113.

Scottish Aquired Brain Injury Network (2017) *Assessment of Cognition – Brain Injury E-Learning Resource*. Available at: https://www.acquiredbraininjury-education.scot.nhs.uk/impact-of-abi/cognitive-problems/assessment-of-cognition/ (Accessed: 13 December 2020).

Scottish Government (2020) *Scottish Index of Multiple Deprivation*. Available at: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/ (Accessed: 10 December 2020).

Spitz, G. *et al.* (2012) 'Association Between Cognitive Performance and Functional Outcome Following Traumatic Brain Injury: A Longitudinal Multilevel Examination', *Neuropsychology*, 26, pp. 604–12. Available at: https://doi.org/10.1037/a0029239.

Strauss, P. of P.E. *et al.* (2006) *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Oxford University Press.

Teager, A. *et al.* (2020) 'The use of cognitive screens within major trauma centres in England: A survey of current practice', *Trauma*, 22(3), pp. 201–207. Available at: https://doi.org/10.1177/1460408619871801.

Teasdale, G. and Jennett, B. (1974) 'Assessment of Coma and Impaired Consciousness: A Practical Scale', *The Lancet*, 304(7872), pp. 81–84. Available at: https://doi.org/10.1016/S0140-6736(74)91639-0.

Theiling, J., Petermann, F. and Daseking, M. (2013) 'WAIS-IV Profiles in First-Ever Unilateral Ischemic Stroke Patients', *Zeitschrift für Neuropsychologie*, 24, pp. 239–252. Available at: https://doi.org/10.1024/1016-264X/a000104.

Wechsler, D. (2008) *Wechsler Adult Intelligence Scale—Fourth Edition: Technical and interpretive manual.* San Antonio, TX: Pearson Assessment.

Weil, Z.M., Corrigan, J.D. and Karelina, K. (2018) 'Alcohol Use Disorder and Traumatic Brain Injury', *Alcohol Research : Current Reviews*, 39(2), pp. 171–180.

Whyte, E. *et al.* (2011) 'Cognitive Impairment in Acquired Brain Injury: A Predictor of Rehabilitation Outcomes and an Opportunity for Novel Interventions', *PM & R : the journal of injury, function, and rehabilitation*, 3(6 0 1), pp. S45–S51. Available at: https://doi.org/10.1016/j.pmrj.2011.05.007.

Wilson, B.A. *et al.* (1996) *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Bury St Edmonds, UK: Thames Valley Test Company.

Wilson, B.A. *et al.* (2008) *Rivermead Behavioural Memory Test*. 3rd edn. Oxford, England: Pearson.

Appendix 1.1 PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported (page)
TITLE			
Title	1	Identify the report as a systematic review.	9
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	10
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	11-12
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	12
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	12-15
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	15 - 16
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Appendix 2.2
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	15-16
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	16
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	13-14, 16; Tables 1.1-1.5
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	As above
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	20
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	13-14, 16; Appendix 1.3
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	19
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	17
Study characteristics	17	Cite each included study and present its characteristics.	Tables 1.1-1.5
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Tables 1.3-1.6
Results of	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision	Tables 1.3-

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported (page)
individual studies		(e.g. confidence/credible interval), ideally using structured tables or plots.	1.5
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	48-49
	23b	Discuss any limitations of the evidence included in the review.	48
	23c	Discuss any limitations of the review processes used.	49
	23d	Discuss implications of the results for practice, policy, and future research.	48-49
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	12
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	12
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	12
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	n/a
Competing interests	26	Declare any competing interests of review authors.	n/a
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	n/a

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

Figure 1. 2 PRISMA Checklist; Page et al., (2021) – Synthesis methods; reporting bias assessment; certainty assessment; results of syntheses; reporting biases; and certainty of evidence not included, as not relevant

Appendix 1.2: Search Strategies

Database searched: APA PSYCINFO (EBSCO Host)

1. DE(Traumatic Brain injury)

OR

2. TX(“Traumatic head injur*” OR “traumatic brain injur*” OR “head injur*” OR “head trauma” OR “brain trauma” OR “concussion” OR “TBI” OR “HI” or “mTBI” or “brain injur*")

AND

3. DE(Cognitive Impairment OR Cognitive Dysfunction OR cognitively impaired OR Neurocognition OR Executive Function OR Memory)

OR

4. TX(“cognit*” OR “neuro cognit*” OR “neuropsycholog*” OR “neuro#cognit*” OR “neuro#psycholog*” OR “executive function*” OR “memory”)

AND

5. DE(Screening Tests or Screening)

OR

6. TX(“screen* tool*” OR “screen* assessment*” OR “screen* measure*” OR “screen* test*” OR “screen* instrument*” OR “screen*” OR “MMSE” OR “Mini Mental State Exam” OR “MoCA” OR “Montreal Cognitive Assessment”)

AND

7. DE(Psychometrics OR Test Validity OR Test Reliability)

OR

8. TX(“psychometric*” OR “valid*” OR “reliab*” OR “sensitiv* OR “specific* OR “receiver operating characteristic” OR “ROC” OR “clinical utility” OR “clinically useful”)

Database searched: CINAHL (EBSCO Host)

1. DH(“Brain Injuries”)

OR

2. TX(“Traumatic head injur*” OR “traumatic brain injur*” OR “head injur*” OR “head trauma” OR “brain trauma” OR “concussion” OR “TBI” OR “HI” or “mTBI” or “brain injur*")

AND

3. DH(“Cognition Disorders” OR “Neurocognitive (Iowa NOC)” OR “Cognition” OR “Executive Function” OR “Memory”)

OR

4. TX(“cognit*” OR “neuro cognit*” OR “neuro psycholog*” OR “neuro#cognit*” OR “neuro#psycholog*” OR “executive function*” OR “memory”)

AND

5. TX(“screen* tool*” OR “screen* assessment*” OR “screen* measure*” OR “screen* test*” OR “screen* instrument*” OR “screen*” OR “MMSE” OR “Mini Mental State Exam” OR “MoCA” OR “Montreal Cognitive Assessment”)

AND

6. DH(Psychometrics OR Reliability OR Validity OR Reliability and Validity)

OR

7. TX(“psychometric*” OR “valid*” OR “reliab*” OR “sensitiv* OR “specific* OR “receiver operating characteristic” OR “ROC” OR “clinical utility” OR “clinically useful”)

Database Searched: Web of Science Core Collection (Web of Science; Clarivate)

1. TX(“Traumatic head injur*” OR “traumatic brain injur*” OR “head injur*” OR “head trauma” OR “brain trauma” OR “concussion” OR “TBI” OR “HI” or “mTBI” or “brain injur*")

AND

2. TX(“cognit*” OR “neuro cognit*” OR “neuro psycholog*” OR “neuro$cognit*” OR “neuro$psycholog*” OR “executive function*” OR “memory”)

AND

3. TX(“screen* tool*” OR “screen* assessment*” OR “screen* measure*” OR “screen* test*” OR “screen* instrument*” OR “screen*” OR “MMSE” OR “Mini Mental State Exam” OR “MoCA” OR “Montreal Cognitive Assessment”)

AND

4. TX(“psychometric*” OR “valid*” OR “reliab*” OR “sensitiv* OR “specific* OR “receiver operating characteristic” OR “ROC” OR “clinical utility” OR “clinically useful”)

https://www.webofscience.com/wos/woscc/summary/2945d620-576a-491e-ae9d-bcf8863db6a1-20f9598c/relevance/1

Database Searched: MEDLINE (OVID)

Searched in OVID MEDLINE and EPUB Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions.

1. MH (Brain Injuries, Traumatic)

OR

2. (traumatic head injur* OR traumatic brain injur* OR head injur* OR head trauma OR brain trauma OR concussion OR TBI OR HI OR mTBI OR brain injur*).mp

AND

3. MH(Cognitive Dysfunction OR Cognition OR Executive Function OR Memory)

OR

4. (cognit* OR neuro cognit* OR neuro psycholog* OR neuro?cognit* OR neuro?psycholog* OR executive function* OR memory).mp

AND

5. MH(Mass Screening)

OR

6. (screen* tool* OR screen* assessment* OR screen* measure* OR screen* test* OR screen* instrument* OR screen* OR MMSE OR Mini Mental State Exam OR MoCA OR Montreal Cognitive Assessment).mp

AND

7. MH(Psychometrics OR Sensitivity and Specificity OR reproducibility of results)

OR

8. (psychometric* OR valid* OR reliab* OR sensitiv* OR specific* OR receiver operating characteristic OR ROC OR clinical utility OR clinically useful).mp

Database Searched: EMBASE (OVID)

1. MH (traumatic brain injury)

OR

2. (Traumatic head injur* OR traumatic brain injur* OR head injur* OR head trauma OR brain trauma OR concussion OR TBI OR HI OR mTBI OR brain injur*).mp

AND

3. MH(Cognitive Defect OR Cognition OR Executive Function OR Memory)

OR

4. (cognit* OR neuro cognit* OR neuro psycholog* OR neuro?cognit* OR neuro?psycholog* OR executive function* OR memory).mp

AND

5. MH(screening test OR screening)

OR

6. (screen* tool* OR screen* assessment* OR screen* measure* OR screen* test* OR screen* instrument* OR screen* OR MMSE OR Mini Mental State Exam OR MoCA OR Montreal Cognitive Assessment).mp

AND

7. MH(psychometry or psychometric screening or validity or reliability OR sensitivity analysis OR "sensitivity and specificity")

OR

8. (psychometric* OR valid* OR reliab* OR sensitiv* OR specific* OR receiver operating characteristic OR ROC OR clinical utility OR clinically useful).mp

Appendix 1.3 Risk of Bias and Measurement Property Rating Procedure (COSMIN)

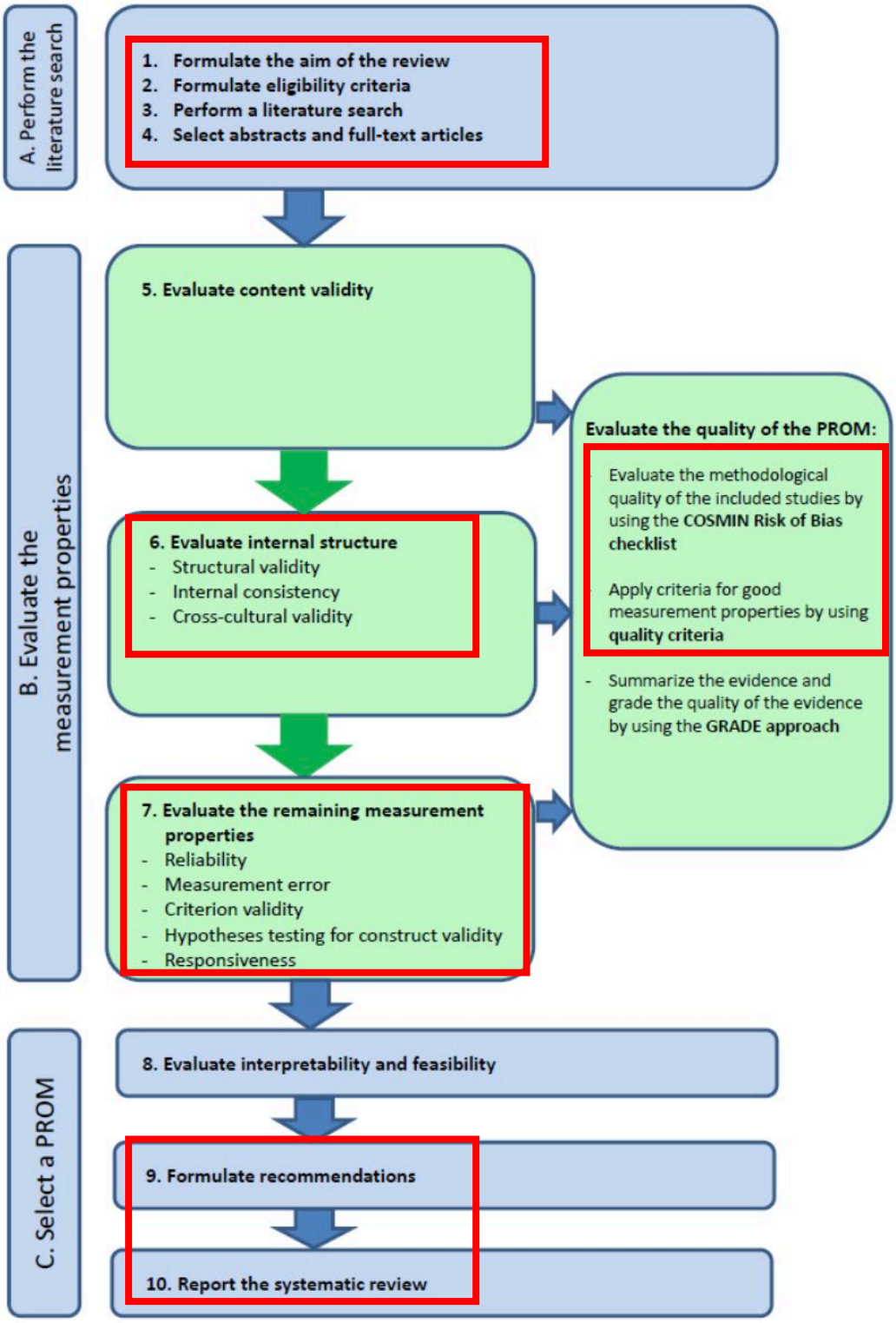


Figure 2. Ten steps for conducting a systematic review of PROMs (2)

Figure 1. 3 Amended diagram from (Mokkink et al., 2010, p19), with steps followed in red

Table 4. Updated criteria for good measurement properties

Measurement property	Rating[1]	Criteria
Structural validity	+	**CTT:** CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08[2] **IRT/Rasch**: No violation of unidimensionality[3]: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 *AND* no violation of local independence: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37 *AND* no violation of monotonicity: adequate looking graphs OR item scalability >0.30 *AND* adequate model fit: IRT: χ^2 >0.01 Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > -2 and <2
	?	CTT: Not all information for '+' reported IRT/Rasch: Model fit not reported
	–	Criteria for '+' not met
Internal consistency	+	At least low evidence[4] for sufficient structural validity[5] AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale[6]
	?	Criteria for "At least low evidence[4] for sufficient structural validity[5]" not met
	–	At least low evidence[4] for sufficient structural validity[5] AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale[6]

Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	–	ICC or weighted Kappa < 0.70
Measurement error	+	SDC or LoA < MIC[5]
	?	MIC not defined
	–	SDC or LoA > MIC[5]
Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis[7]
	?	No hypothesis defined (by the review team)
	–	The result is not in accordance with the hypothesis[7]
Cross-cultural validity\measurement invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's R^2 < 0.02)
	?	No multiple group factor analysis OR DIF analysis performed
	–	Important differences between group factors OR DIF was found
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
	?	Not all information for '+' reported
	–	Correlation with gold standard < 0.70 OR AUC < 0.70
Responsiveness	+	The result is in accordance with the hypothesis[7] OR AUC ≥ 0.70
	?	No hypothesis defined (by the review team)
	–	The result is not in accordance with the hypothesis[7] OR AUC < 0.70

The criteria are based on e.g. Terwee et al.(30) and Prinsen et al.(5)

AUC = area under the curve, CFA = confirmatory factor analysis, CFI = comparative fit index, CTT = classical test theory, DIF = differential item functioning, ICC = intraclass correlation coefficient, IRT = item response theory, LoA = limits of agreement, MIC = minimal important change, RMSEA: Root Mean Square Error of Approximation, SEM = Standard Error of Measurement, SDC = smallest detectable change, SRMR: Standardized Root Mean Residuals, TLI = Tucker-Lewis index

[1] "+" = sufficient, " –" = insufficient, "?" = indeterminate
[2] To rate the quality of the summary score, the factor structures should be equal across studies
[3] unidimensionality refers to a factor analysis per subscale, while structural validity refers to a factor analysis of a (multidimensional) patient-reported outcome measure
[4] As defined by grading the evidence according to the GRADE approach
[5] This evidence may come from different studies
[6] The criteria 'Cronbach alpha < 0.95' was deleted, as this is relevant in the development phase of a PROM and not when evaluating an existing PROM.
[7] The results of all studies should be taken together and it should then be decided if 75% of the results are in accordance with the hypotheses

Figure 1. 4 Summary of measurement property rating guidelines, taken from (Mokkink et al., 2010, pp 28-29)

Adaptations made to COSMIN (based on adaptations used in DClinPsy thesis by Bronagh Reynolds, available online at https://theses.gla.ac.uk/81890/):

The procedure recommended by COSMIN was followed as outlined above. Content validity was not evaluated as it was out-with the scope of this review, and many of the cognitive screens identified were established cognitive screening tools.

As cognitive screening tools can be argued to be unidimensional or multidimensional; tools were taken to be unidimensional unless specified or implied otherwise by a study.

For hypothesis testing designs, it was specified whether it was discriminant or convergent validity that had been assessed. COSMIN's grading of the measurement properties of construct validity (convergent validity) do not specify designs or statistics; their risk of bias rating guidelines advise "correlations between the PROM and the comparator instrument" are an example of an appropriate measure (Mokkink et al., 2010, pp 59). Studies were excluded where their designs did not appear to fit with COSMIN's expectations; for example, multiple regressions where it was difficult to parse out the direct relationship between the tool and the relevant outcome. It was decided that cognitive screens were permissible as a convergent measure and would be rated as "V" (unless contraindicatory information was identified), as (until this review) no review has adequately summarised the psychometric properties of screens in this population.

A discriminant validity column was added to the results table templates provided by COSMIN; to make them consistent with their risk of bias ratings. COSMIN guidance states that p-values are not sufficient for assessing discriminant validity. It was decided that these studies would be rated as "I" inadequate and their measurement properties as (?) insufficient.

For criterion validity, COSMIN specifies "the review team should determine what reasonable 'gold standards' are for the construct to be measured". Any specified TBI diagnostic criteria or diagnostic process was permitted, given the variability in TBI diagnostic procedures internationally; unless poorly defined. Any validated, standard neuropsychological test was permitted. If the study had transformed a continuous variable into a dichotomous one; this was acceptable if they reported appropriate statistics.

COSMIN recommends a Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach to synthesise the evidence across studies. Given the heterogeneity of the literature, it was considered unnecessary for this review.

Appendix 2.0 MRP Proposal

The proposal for this MRP is available at:

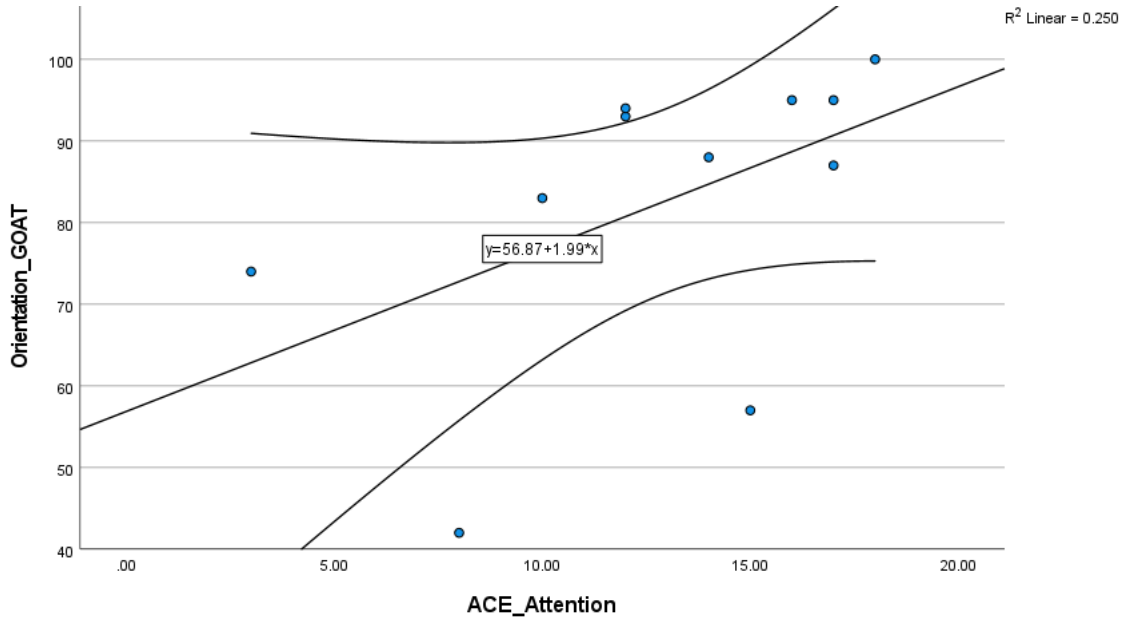
https://osf.io/p4jah/?view_only=70b26c77b05f4fe2a10aed68ae1b03c7

Document title “McLaren_prososal_v3_26-7-21.docx”

Appendix 2.7 Correlation Analyses

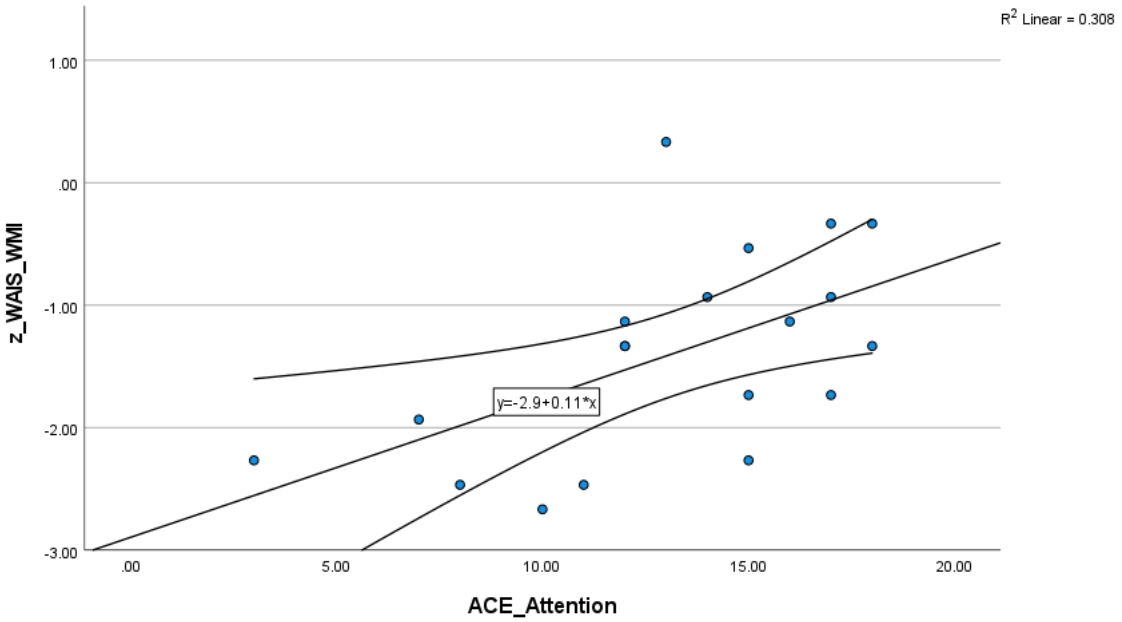
Correlation Analyses. Correlations between TEA and BADS subtests (with associated ACE-III subtests) not included as small sample sizes and ordinal scores made resulting scatterplots less meaningful to interpret.

Figure 2. 1 Scatterplot of GOAT (total score) x ACE-III Attention



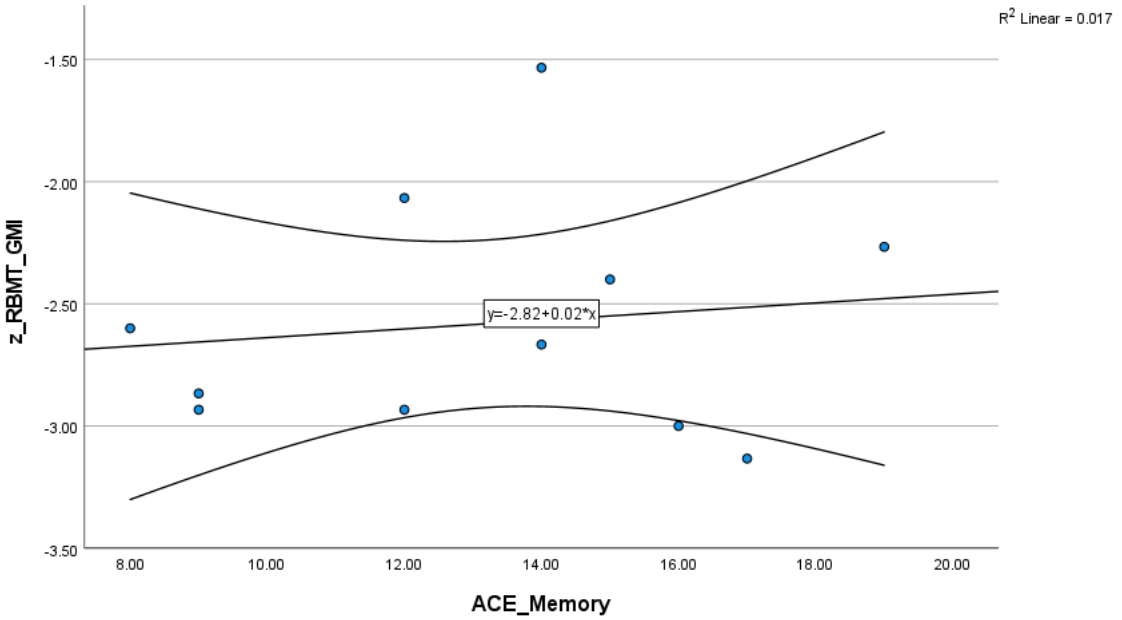
Orientation_GOAT, total score Galveston Orientation Amnesia Test; ACE_Attention, Addenbrookes Cognitive Exam-III Attention. With R^2 fit line and adjacent lines indicating 95% CI.

Figure 2. 2 Scatterplot of WAIS-IV WMI (z score) x ACE-III Attention



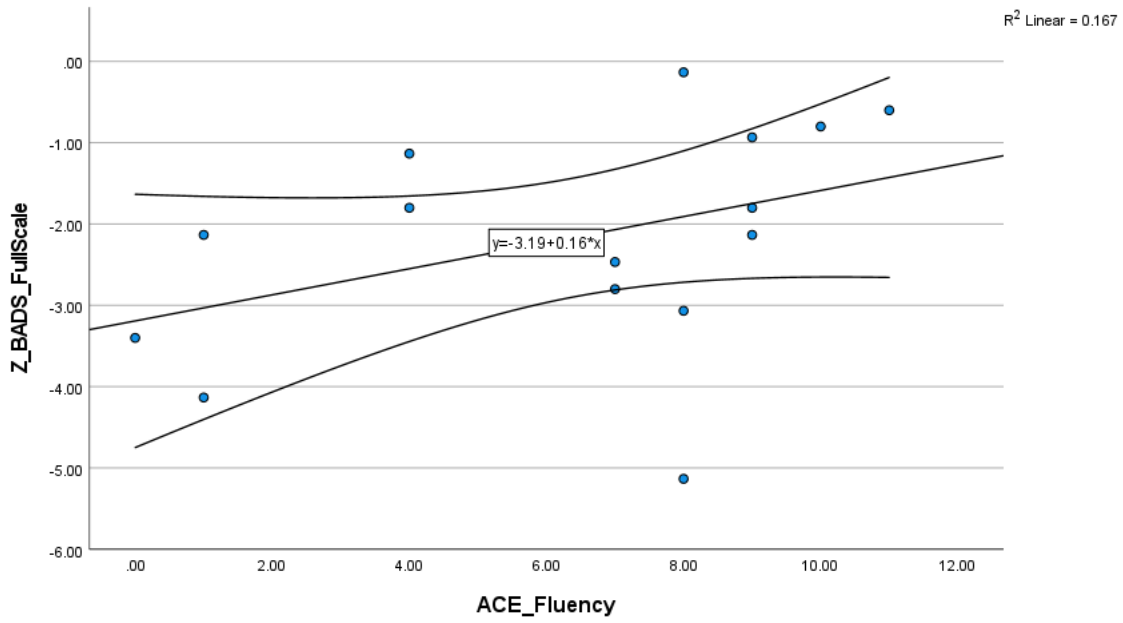
ACE_Attention, Addenbrookes Cognitive Exam-III Attention; z_WAIS_WMI, z score of WAIS, Wechsler Adult Intelligence Scale (WMI, working memory index, z score). With R^2 fit line and adjacent lines indicating 95% CI.

Figure 2. 3 Scatterplot of RBMT (total score; z score) x ACE-III Memory



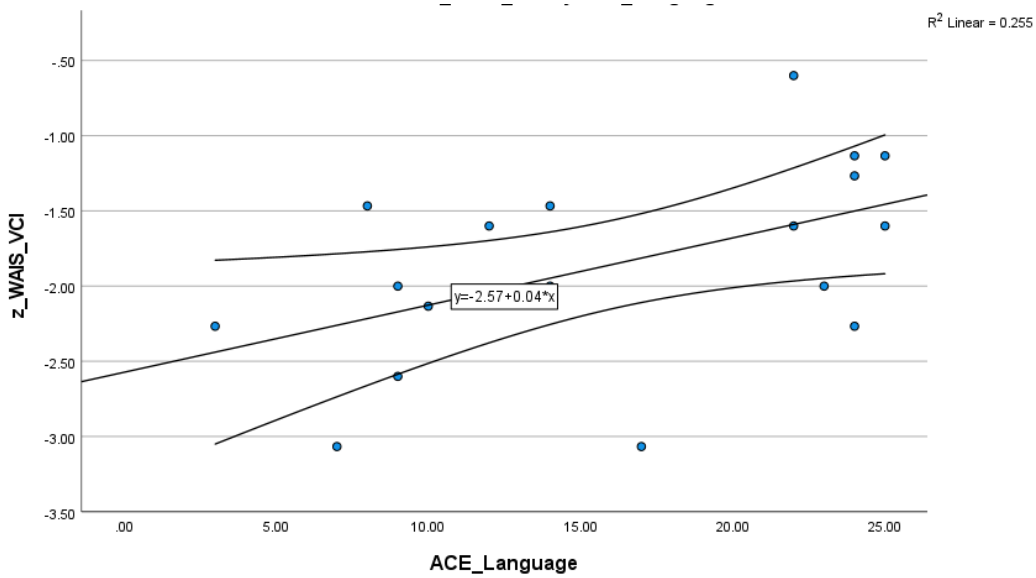
ACE_Memory, Addenbrookes Cognitive Exam-III Memory; z_RBMT_GMI, RBMT Rivermead Behavioural Memory Test General Memory Index (z score); With R^2 fit line and adjacent lines indicating 95% CI.

Figure 2. 4 Scatterplot of BADS x ACE-III Fluency



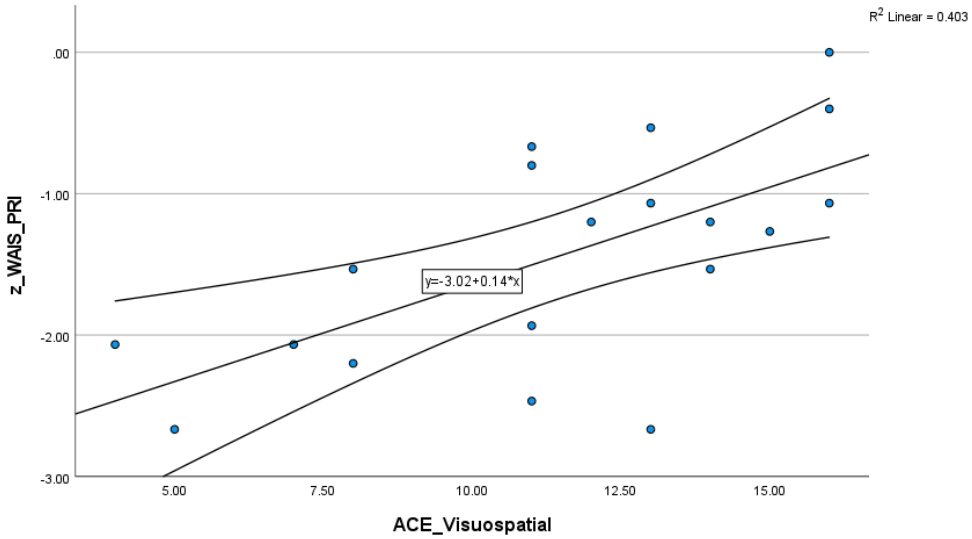
ACE_Fluency, Addenbrookes Cognitive Exam-III fluency; Z_BADS_FullScale, BADS Behavioural Assessment of Dysexecutive Syndrome Full Scale (z score). With R^2 fit line and adjacent lines indicating 95% CI.

Figure 2. 5 Scatterplot of WAIS-IV VCI x ACE-III Language



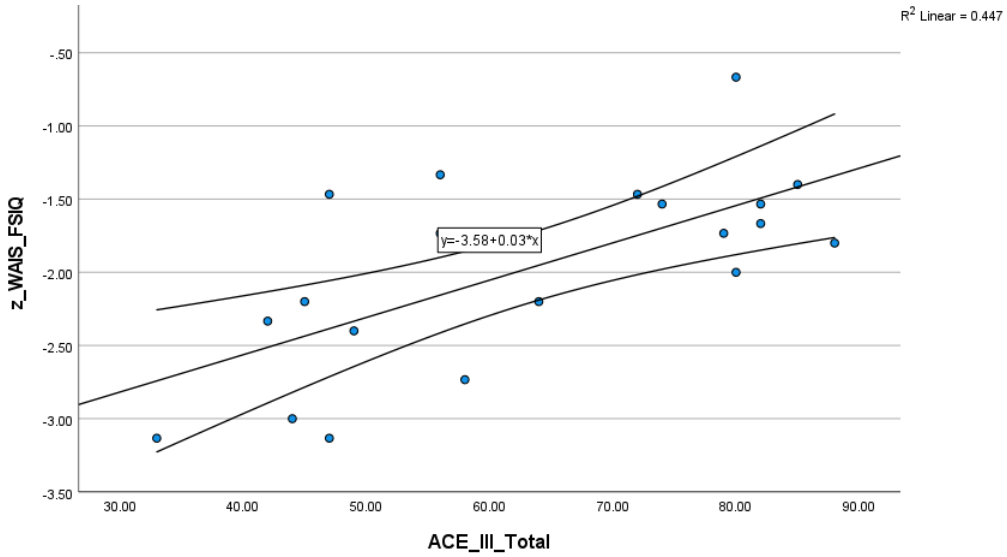
ACE_Language, Addenbrookes Cognitive Exam-III Language; zWAIS_VCI, Wechsler Adult Intelligence Scale (VCI, verbal comprehension index, z score). With R^2 fit line and adjacent lines indicating 95% CI.

Figure 2. 6 Scatterplot of WAIS-IV PRI x ACE-III Visuospatial



ACE_Visuospatial, Addenbrookes Cognitive Exam-III visuospatial; z_WAIS_PRI, Wechsler Adult Intelligence Scale (PRI, perceptual reasoning index, z score). With R^2 fit line and adjacent lines indicating 95% CI.

Figure 2. 7 Scatterplot of WAIS-IV FSIQ x ACE-III Total Score



ACE_III_Total, Addenbrookes Cognitive Exam-III total score; z_WAIS_FSIQ, Wechsler Adult Intelligence Scale (FSIQ, full scale intelligence quotient, z score). With R^2 fit line and adjacent lines indicating 95% CI.

Appendix 2.8 ROC Curves

ROC curve for ACE-III total score and WAIS-IV FSIQ not reported as *"at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased".*

Figure 2. 8 ROC curve of ACE-III Attention and impairment on GOAT

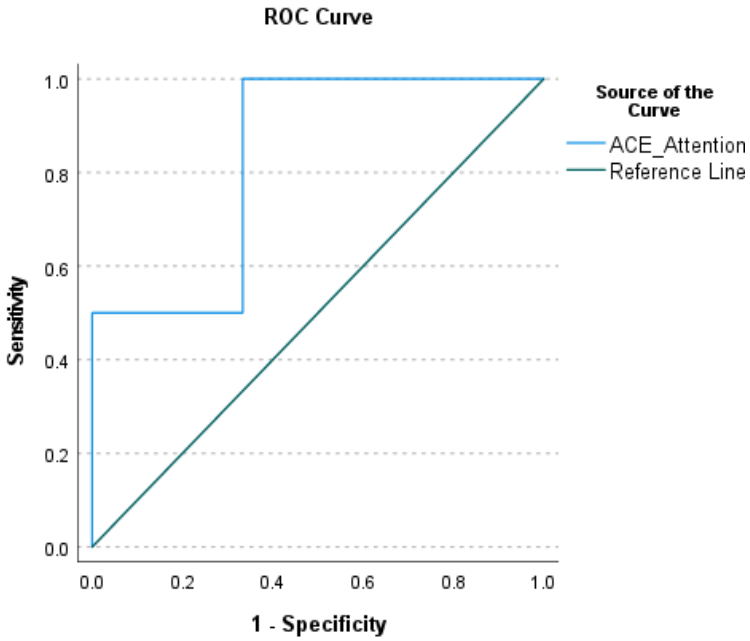


Figure 2. 9 ROC curve of ACE-III Attention and impairment on WAIS-IV WMI

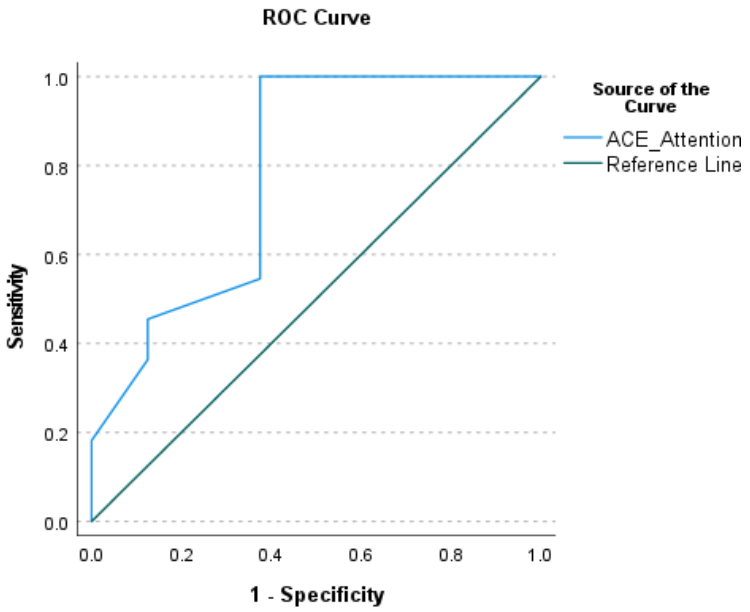


Figure 2. 10 ROC Curve of ACE-III Fluency and impairment on BADS Full Scale

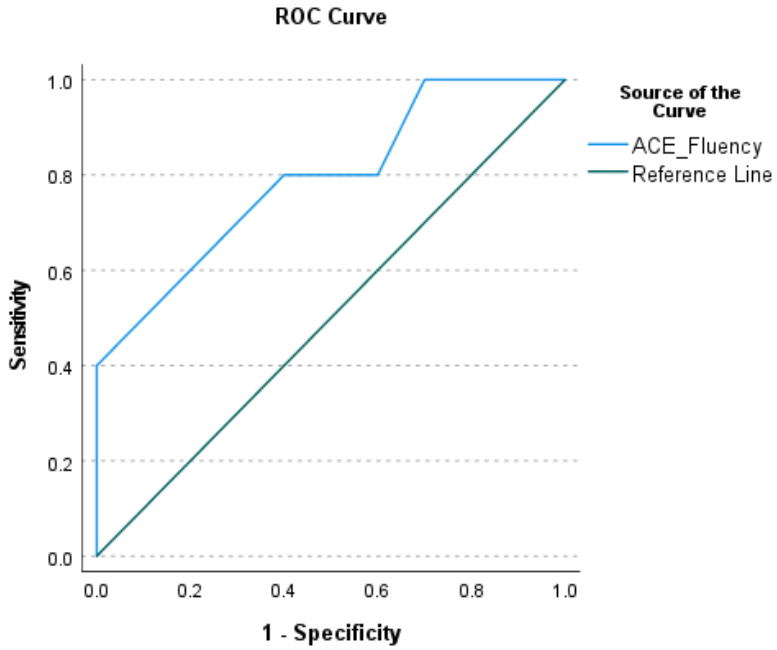


Figure 2. 11 ROC Curve of ACE-III Language and impairment on the WAIS VCI

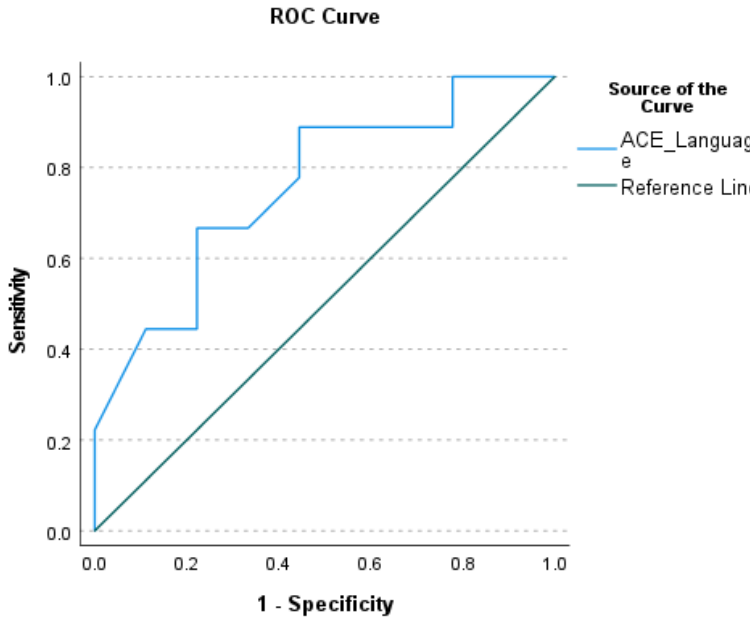


Figure 2. 12 ROC Curve of ACE-III Visuospatial and impairment on the WAIS PRI