An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

and

Clinical Research Portfolio

Hollie Thomson

BA (Hons), MSc

Submitted in part fulfilment of the requirements for the Degree of Doctorate in Clinical Psychology (D.Clin.Psy)

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

February 2020

Acknowledgements

I would firstly like to thank all of the NHS GG&C staff who kindly took the time to participate in my research, as well as all of the Psychologists and Team Leads within Older Adult Community Mental Health Teams who assisted with recruitment, without whom this project would not have been possible.

A huge thank you to my research supervisor, Professor Jon Evans, for his invaluable help, reassurance, and guidance throughout. Thanks also to my field supervisor, Dr Stephanie Crawford, for her advice and support, and to Dr Helen Pulford and Dr Julia Cook for their helpful contributions at the beginning of this research project.

Thank you to all of my wonderful family (grandparents, aunties, uncles, cousins – you know who you are!) and friends who have provided continuous enthusiasm, support, and encouragement throughout my years of studying. Special thanks to my parents, Helen and Thomas, for always believing in me, for their valued advice, and for their endless love and support. I could not have done it without you! And to my fiancé Alex, for his love, support, and friendship - I look forward to our life together! I love you all very much.

~ For my grandparents ~

Gerard Campbell

Ellen Campbell

Margaret Thomson

Tommy Thomson

Table of Contents

CHAPTER 1: Systematic Review

Diagnostic Accuracy of the Montreal Cognitive Assessment (MoCA) for Detecting Mild Cognitive Impairment (MCI) and Dementia in Parkinson's Disease: A Systematic Review

Hollie Thomson[1]*

[1]Mental Health and Wellbeing, Institute of Health and Wellbeing, University *of* Glasgow

*Address for Correspondence:
University *of* Glasgow
Mental Health and Wellbeing
Gartnavel Royal Hospital
1st Floor, Administration Building
1055 Great Western Road
Glasgow
G12 0XH
Email: Hollie.Thomson@ggc.scot.nhs.uk

Prepared in accordance with requirements for submission to the International Journal of Geriatric Psychiatry (Appendix 1.1)

Submitted in part fulfilment of the requirements for the Degree of Doctorate in Clinical Psychology (D.Clin.Psy)

Word Count: 8,229

Abstract

Objectives: This systematic review examines the diagnostic accuracy of English and non-English versions of the Montreal Cognitive Assessment (MoCA) for the detection of Dementia and Mild Cognitive Impairment (MCI) in Parkinson's disease. **Methods:** A systematic search of relevant databases including PsycINFO, MEDLINE, EMBASE, and CINAHL was conducted from 2005 to December 2019. Studies which fulfilled the inclusion criteria were reviewed using Standards for Reporting of Diagnostic Accuracy (STARD) guidance. **Results:** Twelve studies were included for review and were assessed as medium to low quality. Studies generally reported good sensitivity for detecting MCI and dementia across different language versions of the MoCA however specificities were low, particularly for MCI, reducing diagnostic accuracy. Optimal cut-off scores also varied, and findings were compromised by methodological limitations. **Conclusion:** The evidence base exploring the diagnostic accuracy of different language versions of the MoCA for individuals with Parkinson's disease is not well established and further research is required. Future research should address the methodological limitations highlighted by this review through adherence to STARD guidelines.

Introduction

Background

Parkinson's disease (PD) is a progressive, neurodegenerative disorder with a global prevalence of 0.5% to 4% in older adults aged 65 years and over (De Lau & Breteler, 2006). It is characterised by both motor symptoms (i.e. tremor, rigidity, bradykinesia, and postural instability) and non-motor symptoms (i.e. mood disturbance, fatigue, autonomic problems, and cognitive dysfunction) (Shulman, De Jager, & Feany, 2011). Cognitive dysfunction is common in PD and can range from mild cognitive impairment (MCI) to dementia. Dementia is an umbrella term for a range of progressive conditions that are characterised by global cognitive decline and which significantly impact function and behaviour (Alzheimer's Society, 2018a). Mild cognitive impairment refers to cognitive decline which is more significant than what would be expected as part of the normal ageing process but is not as severe as dementia and does not significantly impact on functioning (Alzheimer's Society, 2018b). Although criteria for defining MCI and dementia in the general population have been proposed, the Movement Disorder Society Task Force (MDS-TF) developed specific 'gold standard' clinical criteria for diagnosing MCI and dementia in Parkinson's disease (Litvan *et al*., 2012). Research suggests that PD-MCI predicts the development of dementia, which can occur in up to 80% of people with PD (Aarsland *et al*., 2003) thus identifying PD-MCI early on is a clinical imperative. PD-D can have a substantial impact on functioning, and is associated with reduced quality of life (Litvan *et al.,* 2011), higher risk of institutionalisation (Emre *et al.,* 2007) increased caregiver burden (Aarsland *et al.,* 1999), and higher mortality rates (Levy *et al.,* 2002). The early detection of cognitive dysfunction is therefore essential to ensuring early intervention and optimal management, and improving quality of life (NICE, 2006).

Cognitive Screening Tests

Given that the use of neuropsychological testing in routine clinical practice is limited due to time pressures and resource constraints, there is a need for a brief, simple, as well as accurate (i.e. reliable and valid), cognitive screening test that is acceptable to patients which can be used as part of the initial stage of a comprehensive dementia assessment (Cullen *et al*., 2007). The main aim of a cognitive screening test is to provide information about the presence, or absence, of cognitive impairment based on a person's score on the test compared to referenced norms (Cullen *et al*., 2007). Most screening tests utilise a cut-off score to establish the point at which a person's score moves from being within the 'normal' range to

being within the 'clinical' range, i.e. indicating the presence of a cognitive impairment. The psychometric robustness of a cognitive screening test is crucial to achieving this aim (Cullen *et al*., 2007).

Two key aspects of test validity are sensitivity and specificity, and are concerned with the accuracy of a screening test relative to a reference standard. Sensitivity refers to the proportion of individuals who have the target condition (reference standard positive) and who have a positive test result, and specificity refers to the proportion of individuals who do not have the target condition (reference standard negative) and who have a negative test result (Florkowski, 2008). Thus, sensitivity may be defined as the extent to which a cognitive screening test correctly identifies all those with cognitive impairment, and specificity may be defined as the extent to which a cognitive screening test correctly identifies all those without cognitive impairment. The sensitivity and specificity of a screening test will vary depending on the selected cut-off score; as the cut-off score of a test is manipulated to maximise sensitivity, the specificity of the test decreases and vice versa (Gifford and Cummings, 1999). This inverse relationship between sensitivity and specificity is illustrated by a Receiver Operating Characteristic (ROC) curve. A ROC curve is a plot of the true positive rate (i.e. sensitivity) against the false positive rate (i.e. specificity) for the different cut-off points of a screening test and can therefore determine the optimal cut-off point for clinical use (Florkowski, 2008). A perfect test will produce an area under the ROC curve of 1 (Gifford and Cummings, 1999). A cognitive screening test with good sensitivity and specificity increases its clinical utility in that all those with a cognitive impairment are accurately classified as such, and that those without a cognitive impairment are not misclassified as impaired and referred on for further testing unnecessarily (Stolwyk *et al*., 2014). It should be noted however that in order to accurately assess the test validity of a screening measure, there must be a 'gold standard' test available for comparison.

An increasing number of cognitive screening tests are being translated for use in multiple languages and across cultures. It should be noted however that the clinical utility of a test will be limited without consideration of its reliability and validity in the country in which it is being used. The psychometric properties of a cognitive screening test may vary between language versions due to differences in populations across studies such as age, education and cultural factors (Lewis *et al*., 2009; O'Driscoll & Shaikh, 2017). For example, the quality of education varies across countries thus years of education may not be equivalent. Wong et al. (2009) pointed out that many older adults in Asian countries have received

significantly less education than their Western counterparts. Poor translations, a lack of cultural equivalents, or words that are out with the general understanding or vocabulary of particular cultures may also affect the validity of a translated test. Consequently, individuals from different countries and cultural backgrounds are unlikely to perform consistently on a cognitive screening test due to factors other than cognitive decline.

The MoCA

The Montreal Cognitive Assessment (MoCA; Nasreddine *et al.,* 2005) is a brief cognitive screening test which takes approximately 10 minutes to administer, and assesses several cognitive domains including orientation, attention, memory, language, visuospatial function, and executive function (Nasreddine *et al.,* 2005). A maximum score of 30 can be obtained, with higher scores indicative of better cognitive performance. A cut-off score of 26 or above is considered to be within the 'normal' range (Nasreddine *et al.,* 2005). The MoCA is easily accessed and is available in different language versions. Although the MoCA was not developed specifically for patients with PD, it has been found to be useful in identifying individuals with MCI and dementia in non-PD populations (Larner, 2012). The aim of the current paper therefore is to review the evidence in relation to the diagnostic accuracy of English and non-English versions of the MoCA in diagnosing MCI and/or dementia in patients with PD.

Methods

Search Strategy

The following electronic bibliographic databases were searched from 2005 (year of publication of original MoCA paper) to December 2019: MEDLINE (OVID), EMBASE (OVID), PsychINFO (EBSCO), and CINAHL (EBSCO). The following combinations of search terms were used in all databases: (["Montreal Cognitive Assessment" OR "MoCA"] AND ["dementia" OR "Alzheimer's" OR "cognitive impairment"] AND ["Parkinson* disease" OR "PD"]). Given that Alzheimer's is the most common form of dementia, it was included in the search terms. All searches were limited to English language. Titles and abstracts of studies generated by initial searches were screened for relevance, and potentially eligible studies were reviewed in full against the inclusion criteria. Reference lists of all included papers were also examined to identify any further relevant studies.

Inclusion and Exclusion Criteria

The inclusion criteria were:

1. Studies investigating the diagnostic accuracy of the MoCA for detecting mild cognitive impairment (MCI) and/or dementia in patients with Parkinson's disease.
2. Studies investigating the diagnostic accuracy of both English and non-English versions of the MoCA.

The exclusion criteria were:

1. Studies that were not in English.
2. Studies that did not use the Movement Disorder Society Task Force (MDS-TS) criteria for diagnosing MCI or dementia.
3. Studies that used the MoCA to track changes in cognitive functioning over time rather than assessing diagnostic accuracy.
4. Studies that used the MoCA as part of a wider cognitive assessment without providing information on diagnostic accuracy.
5. Studies investigating the Short Form version of the MoCA (SF-MoCA) or the MoCA Basic (MoCA-B).
6. Abstracts, response letters, reviews, and guides.

Methodological Quality

In order to rate the methodological quality of the studies included in this review, the study quality was critically appraised and scored by the author using the Standards for the Reporting of Diagnostic accuracy studies guidelines (STARD; Bossuyt *et al*., 2015; Appendix 1.2). The STARD provides international consensus guidelines for evaluating diagnostic accuracy studies. It consists of 31 items in total. Each item has a maximum score of 2 points indicating that information is present, with 1 point indicating that information is present but with insufficient details, and 0 points indicating that information is missing. The STARD therefore has a maximum score of 62 points. Scores were summed to provide an overall quality score for each of the studies included. An independent rater assessed 50% of the studies to certify that assessment scores were reliable. There was 92% agreement between raters, and where discrepancies occurred, consensus was reached through discussion. Overall scores were then combined with an assessment of the extent to which the checklist criteria had been fulfilled to provide an overall quality rating of high (++), medium (+), or low (–) (NICE, 2014).

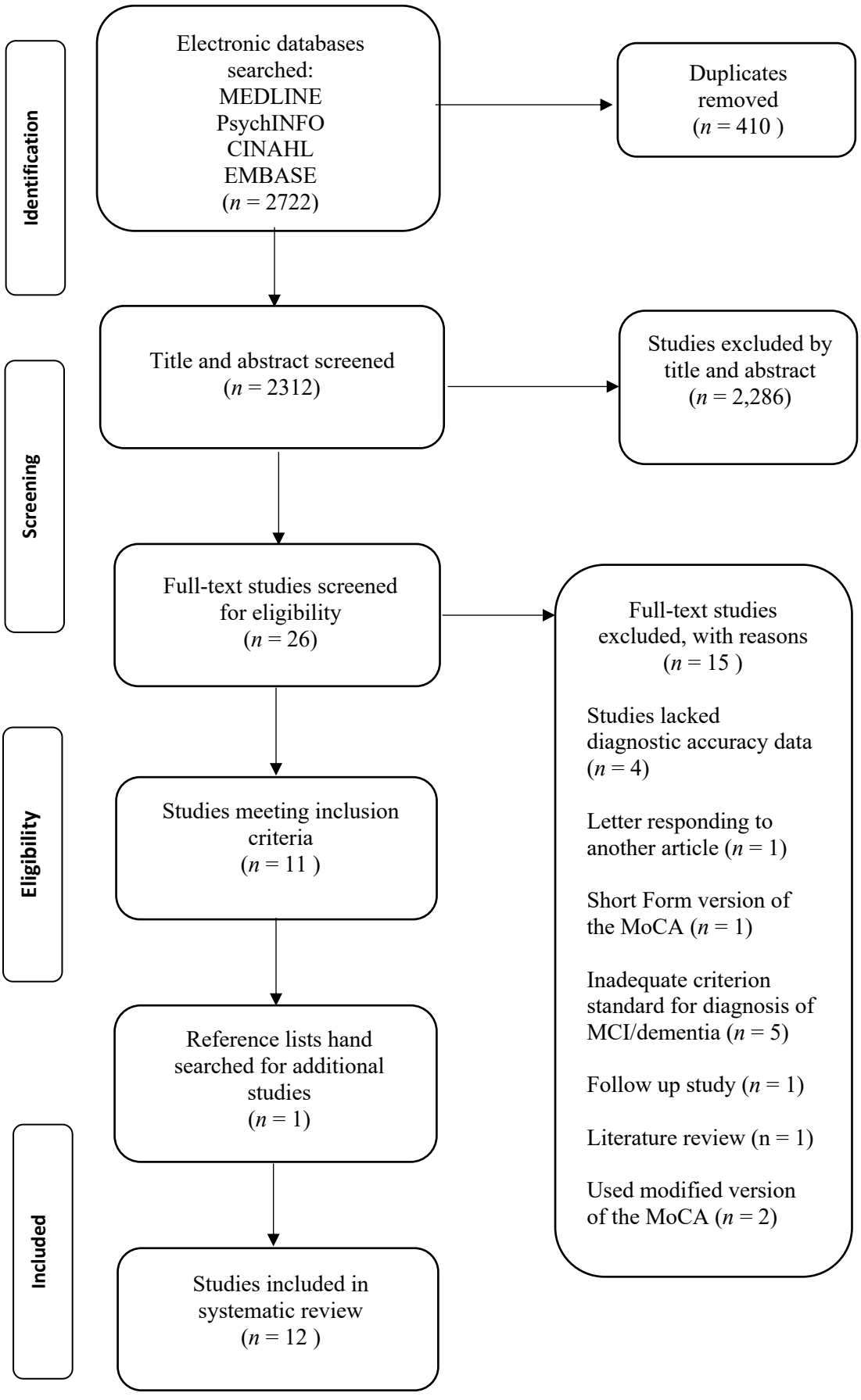

Figure 1 PRISMA flow diagram illustrating the search process

Results

Outcome of Search Process

A PRISMA flow diagram of search results is displayed in Figure 1 depicting the article search and review process. The initial search yielded 2722 studies of which 410 were duplicates. Titles and abstracts were subsequently reviewed by the author using the study selection criteria outlined above which resulted in 26 remaining studies. Upon full-text review of the remaining articles, 12 studies were included for review.

Study Characteristics

A summary of the study characteristics of all eligible studies is reported in Table 1. The studies reported on English, Portuguese, Italian, Chinese, and Turkish versions of the MoCA. The mean age of participants with Parkinson's disease and either MCI or dementia in the studies ranged between 57 (Sobreira *et al.*, 2015) and 73.4 (Dalrymple-Alford *et al*., 2010). Years of education ranged between 5.5 (Sobreira *et al.*, 2015) and 16.2 (Hoops *et al*., 2009), and the male to female ratio varied between 16:15 (Sobreira *et al.*, 2015) and 299:96 (Hendershott *et al.*, 2019).

Information about diagnostic accuracy, including cut-off scores, sensitivity, and specificity, for each of the studies are reported in Table 2. The cut-off scores for the different language versions of the MoCA included in this review ranged between 22.5 (Almeida *et al.,* 2019) and 26 (Marras *et al.*, 2013; Kandiah *et al.*, 2014; Sobreira *et al.*, 2015; Hendershott *et al.*, 2019) for MCI, and between 17.5 (Almeida *et al.,* 2019) and 24.5 (Hoops *et al*., 2009) for dementia. The sensitivity of the measures ranged between 77% (Brown *et al.*, 2016) and 93% (Kandiah *et al.*, 2014) for MCI, and between 59% (Ozdilek *et al.*, 2014) and 94% (Sobreira *et al.*, 2015) for dementia. Specificity values ranged between 27% (Sobreira *et al.*, 2015) and 79% (Brown *et al.*, 2016) for MCI, and between 76% (Almeida *et al.,* 2019) and 89% (Ozdilek *et al.*, 2014; Camargo *et al.*, 2016) for dementia.

Table 1 Study Characteristics

Study	Language	Participant type	Gender male:female	Age in years Mean ± SD	Years of education Mean ± SD	MoCA Mean ± SD
Almeida *et al.* (2019)	Portuguese (Brazilian)	PD-D (*n* = 25)	11:14	65.48 ± 8.67	5.63 ± 4.68	15.28 ± 3.6
		PD-MCI (*n* = 38)	18:20	66.53 ± 8.64	8.74 ± 5.54	20.76 ± 3.67
		PD-NC (*n* = 26)	19:7	57.15 ± 7.81	11.15 ± 4.7	25.42 ± 3.13
Brown *et al.* (2016)	English	PD-MCI (*n* = 58)	26:32	67.43 ± 10.06	15.17 ± 2.15	21.90 ± 3.55
		PD-NC(*n* = 53)	18:35	62.73 ± 8.61	15.98 ± 1.43	25.74 ± 2.57
Camargo *et al.* (2016)	Portuguese (Brazilian)	PD-D (*n* = 41)	26:15	69.78 ± 10.88	5.925 ± 4.99	14.66 ± 5.43
		PD-NC (*n* = 9)	6:3	66.89 ± 14.21	19.44 ± 6.38	22.44 ± 3.36
Chen *et al.* (2013)	Chinese	PD-D (*n* = 235)	112:123	69.3 ± 10.0	9 (6 - 12)*	18 (15 - 21)*
		Healthy (*n* = 381)	263:118	64.8 ± 10.6	12 (9 - 15)*	26 (24 - 28)*
Dalrymple-Alford *et al*. (2010)	English	PD-D (*n* = 21)	18:3	73.4 ± 6.7	12.9 ± 3.0	16.9 ± 4.0
		PD-MCI (*n* = 21)	15:6	71.5 ± 5.4	12.3 ± 3.1	23.2 ± 2.5
		PD-NC (*n* = 72)	50:22	64.5 ± 8.4	13.2 ± 3.0	26.7 ± 2.1
		Healthy (*n* = 47)	31:16	67.3 ± 9.3	13.7 ± 3.0	27.2 ± 1.9
Federico *et al.* (2015)	Italian	PD-MCI (*n* = 22)	15:7	68.9 ± 7.2	8.3 ± 2.8	
		PD-NC (*n* = 21)	12:9	67.5 ± 11.2	8.7 ± 3.1	
Hendershott *et al.* (2019)	English	PD-MCI or PD-D (*n* = 395)	299:96	67.8 ± 8.6	15.6 ± 2.6	23.7 ± 3.2
		PD-NC (*n* = 126)	59:67	63.4 ± 8.7	16.2 ± 2.5	27.4 ± 2.1
Hoops *et al*. (2009)	English	PD-MCI or PD-D (*n* = 40)	33:7	68.1 ± 9.2	16.2 ± 3.1	22.2 ± 4.1
		PD-NC (*n* = 92)	67:25	63.9 ± 9.7	16.5 ± 3.1	26.2 ± 2.9
Kandiah *et al.* (2014)	Chinese	PD-MCI (*n* = 34)	20:14	70.49 ± 6.54	9.29 ± 3.14	23.35 ± 3.03
		PD-NC (*n* = 61)	47:14	64.34 ± 7.73	10.97 ± 3.30	27.66 ± 1.94
Marras *et al.* (2013)	English	PD-MCI (*n* = 46)	29:17	71.1 ± 4.8	15.3 ± 2.7	23.8 ± 3.3
		PD-NC (*n* = 93)	64:29	71.1 ± 5.7	16.1 ± 2.4	25.9 ± 2.4
Ozdilek *et al.* (2014)	Turkish	PD-D (*n* = 9)	6:3	67.4 ± 7.8	5.6 ± 2.0	18.6 ± 3.7
		PD-MCI (*n* = 13)	10:3	63.3 ± 9.3	7.3 ± 3.0	20.6 ± 3.9
		PD-NC (*n* = 28)	18:10	58.3 9.5	10.0 ± 4.2	24.6 ± 3.9
		Healthy (*n* = 50)	22:28	62.3 9.4	10.0 ± 4.2	23.7 ± 4.1
Sobreira *et al.* (2015)	Portuguese (Brazilian)	PD-D (*n* = 16)	3:13	72.5 (53 - 81)*	5.5 (2 - 18)*	17 (7 - 24)*
		PD-MCI (*n* = 31)	16:15	57 (37 - 77)*	10 (0 - 20)*	23 (14 - 29)*
		PD-NC (*n* = 30)	10:20	61 (28 – 79)*	4 (1 - 20)*	23.5 (15 - 29)*

Note: A blank space indicates that no information is available.
Abbreviations: PD, Parkinson's disease; PD-D, PD patient with dementia; PD-MCI, PD patient with mild cognitive impairment; PD-NC, PD patient with normal cognition.
*Only the median (min-max) was reported in the article.

Table 2 Diagnostic accuracy information

Study	Target condition	Cut-off score	Sensitivity[a]	Specificity[b]	AUC	PPV[c]	NPV[d]
Almeida *et al.* (2019)	Dementia	17.5	0.82	0.76			
	MCI	22.5	0.85	0.71			
Brown *et al.* (2016)	MCI	25	0.77	0.79	0.82		
Camargo *et al.* (2016)	Dementia	19	0.88	0.89	0.91	0.97	0.61
Chen *et al.* (2013)	Dementia	23	0.70	0.77	0.83	0.59	0.85
Dalrymple-Alford *et al*. (2010)	Dementia	21	0.81	0.95	0.97	0.87	0.92
Federico *et al.* (2015)	MCI	25.5	0.82	0.67	0.79		
Hendershott *et al.* (2019)	MCI	26	0.80	0.68	0.88	0.89	0.52
Hoops *et al*. (2009)	Dementia	24.5	0.82	0.75	0.87	0.38	0.96
Kandiah *et al.* (2014)	MCI	26	0.93	0.59	0.91		
Marras *et al.* (2013)	MCI	26	0.83	0.44	0.71	0.42	0.83
Ozdilek *et al.* (2014)	MCI or Dementia	21	0.59	0.89	0.79	0.81	0.73
Sobreira *et al.* (2015)	MCI	26	0.84	0.27	0.50		
	Dementia	21	0.94	0.68	0.86		

Note: A blank space indicates that no information is available.
Abbreviations: MCI, Mild Cognitive Impairment; AUC, Area Under the Curve; PPV, Positive Predictive Value; NPV, Negative Predictive Value.
[a]Sensitivity means that the test correctly identified those with MCI or dementia (i.e. the proportion of true positives).
[b]Specificity means that the test correctly identified those without MCI or dementia (i.e. the proportion of true negatives).
[c]PPV means the proportion of people identified as having MCI or dementia who actually do have the condition.
[d]NPV means the proportion of people identified as not having MCI or dementia who do not have the condition.

Table 3 Quality assessment of included studies

Study	STARD score/ number of items	Main limitations	Rating of overall quality
Almeida *et al.* (2019)	34/62	No information on indeterminate or missing data, time interval not reported, non-blinded, small sample, a priori cut-off point for index test not reported, no participant flow chart	-
Brown *et al.* (2016)	35/62	Power not calculated, non-blinded, no information on indeterminate or missing data, no participant flow chart, time interval not reported, a priori cut-off point for index test not reported, no study limitations reported	-
Camargo *et al.* (2016)	35/62	Small sample, non-blinded, no information on indeterminate or missing data, power not calculated, time interval not reported, poorly defined sample, no participant flow chart, a priori cut-off point for index test not reported	-
Chen *et al.* (2013)	42/62	A priori cut-off point for index test not reported, non-blinded, power not calculated, no information on indeterminate or missing data	+
Dalrymple-Alford *et al*. (2010)	42/62	Non-blinded, no information on indeterminate or missing data, power not calculated, a priori cut-off point for index test not reported	+
Federico *et al.* (2015)	37/62	Study timescale not reported, power not calculated, small sample, no information on indeterminate or missing data, a priori cut-off point for index test not reported, time interval not reported	-
Hendershott *et al.* (2019)	32/62	Non-blinded, no information on indeterminate or missing data, power not calculated, a priori cut-off point for index test not reported, time interval not reported, poorly defined sample, no study limitations reported, no participant flow chart	-
Hoops *et al*. (2009)	42/62	Power not calculated, non-blinded, no information on missing data, a priori cut-off point for index test not reported, small sample	+
Kandiah *et al.* (2014)	36/62	Poorly defined sample, small sample, no participant flow chart, no information on indeterminate or missing data, non-blinded, power not calculated, time interval not reported, a priori cut-off point for index test not reported	-
Marras *et al.* (2013)	42/62	No information on indeterminate data, power not calculated, no participant flow chart, a priori cut-off point for index test not reported, no study limitations reported	+
Ozdilek *et al.* (2014)	38/62	No participant flow chart, a priori cut-off point for index test not reported, no information on indeterminate or missing data, power not calculated, poorly defined sample, non-blinded, small sample	-
Sobreira *et al.* (2015)	36/62	Non-blinded, a priori cut-off point for index test not reported, no information on indeterminate or missing data, power not calculated, no study limitations reported, small sample, no participant flow chart	-

Quality Assessment

Of the 12 studies included in this review, eight were assessed as low (–) quality and four were assessed as medium (+) quality (see Table 3). Across all studies, an adequate description of the demographic and clinical characteristics of study participants was reported including age, gender, years of education, and Parkinson's disease duration. All but one study (Almeida *et al.,* 2019) reported Parkinson's disease severity using either the Hoehn and Yahr Scale or the Unified Parkinson's Disease Rating Scale, Part three (UPDRS-III). This information is important given that the demographic and clinical characteristics of a sample can influence the diagnostic accuracy of a test and the generalisability of results (Bossuyt *et al*., 2015). As per the inclusion criteria, all studies utilised the 'gold standard' Movement Disorder Society Task Force (MDS-TF) criteria as a reference standard to establish the presence or absence of MCI or dementia in study participants. These criteria increase the validity of those diagnoses within this particular population (Litvan *et al*., 2012). All studies reported estimates of diagnostic accuracy including sensitivity and specificity, however, other measures of diagnostic accuracy, including positive and negative predictive values, were only reported in seven of the studies. All studies provided an adequate summation of the consequences of their findings and implications for clinical practice.

There were a number of issues common to all or most of the studies. Firstly, all of the studies, with the exception of two (Federico *et al*., 2015; Marras *et al*., 2013), failed to report on whether clinical diagnosis information and reference standard results were available to the assessors of the index test (i.e. the MoCA). Assessors' interpretation of MoCA results may have been influenced by their knowledge of the results of the reference standard thereby introducing review bias (Whiting *et al*., 2004). Indeed, non-blinding of assessors could have resulted in more study participants being accurately diagnosed with MCI or dementia in the studies than would be the case in clinical practice. Secondly, half of the studies failed to report the time interval between administration of the MoCA and the reference standard. If the index test and reference standard are not performed at the same time, changes may occur in the target condition and/or other conditions during the interval period that could lead to biased estimates of test performance (Knottnerus & Muris, 2003). Thirdly, no studies specified a priori defined cut-off scores prior to performing the MoCA, instead choosing the cut-off score which maximised test performance in their sample. This can lead to estimates of sensitivity and specificity that are overly optimistic, particular in studies with small sample sizes. When possible, studies should use prespecified cut-off points to improve the

validity of diagnostic accuracy findings (Leeflang *et al*., 2008). Fourthly, there was little reporting of missing or indeterminate data across all of the studies which could also introduce biased estimates of diagnostic accuracy (Bossuyt *et al*., 2015). And lastly, all of the studies with the exception of one (Almeida *et al.,* 2019) failed to report on power calculation and intended sample size which has implications for the clinical relevance of study findings. Sample size in most of the studies tended to be small which increases the likelihood of imprecise estimates of diagnostic accuracy (Bossuyt *et al*., 2015).

All or most of the issues discussed above were relevant to the studies rated as low quality. It should be noted however that some of these issues may represent poor reporting quality rather than methodological flaws. Nevertheless, these are items that the STARD guidelines consider to be essential and indeed their absence compromises the ability to draw valid conclusions about diagnostic accuracy. Where these issues were given consideration, and at least partially addressed, studies were rated as medium quality. Quality assessment scores for each of the items on the STARD for all included studies can be found in Appendix 1.3. Individual studies are described below and are categorised by language version of the MoCA.

Review of Study Findings

Italian

One study evaluated the diagnostic accuracy of the Italian version of the MoCA in diagnosing MCI in Parkinson's disease (Federico *et al*., 2015). The results suggested that the MoCA could distinguish those with MCI with a sensitivity of 82%, and a specificity of 67%, with a cut-off of 25.5. The MoCA could not however reach a combined sensitivity and specificity of 80% at any cut-off value. The validity of this study for drawing conclusions about the sensitivity and specificity of the Italian version of the MoCA in PD-MCI is limited due to the small number of study participants (n = 43), the high prevalence of MCI included in the sample, and the strict selection criteria. The study also failed to report the time interval between the reference standard and the index test which could have impacted on estimates of diagnostic accuracy. Although the researchers in this study were blind to the clinical status of participants therefore enhancing study validity, the study was rated as low quality overall due to the aforementioned issues which reduce the generalisability of results, and compromise the ability to draw valid conclusions about diagnostic accuracy.

Turkish

One study evaluated the validity of the Turkish version of the MoCA as a screening tool for detecting MCI and dementia in Parkinson's disease (Ozdilek *et al.,* 2014). The results suggested that the optimal MoCA cut-off score for detecting *any* cognitive dysfunction was 21 (sensitivity = 59%, specificity = 89%) however an appropriate cut-off score for differential diagnosis of MCI or dementia was not possible. Interpretation is complicated here by the combined analysis of MCI and dementia. The discriminant validity and diagnostic accuracy of the Turkish version of the MoCA is therefore difficult to assess. The small sample size (n = 50) in this study may also have compromised estimates of diagnostic accuracy. This study was rated as low quality due to the issues discussed above, which compromise the validity of the conclusions that can be made as to diagnostic accuracy.

Chinese

In two studies, the authors investigated the diagnostic accuracy of the Chinese version of the MoCA for detecting MCI and dementia in Parkinson's disease. The studies confirmed the usefulness of the MoCA as a screening tool in Parkinson's disease for both MCI (sensitivity = 93%, specificity = 59%) at a cut-off of 26 (Kandiah *et al.*, 2014), and dementia (sensitivity = 70%, specificity = 77%) at a cut off of 23 (Chen *et al*., 2013) although specificities are fairly low compromising diagnostic accuracy. In any case, the validity of both studies for drawing conclusions about the sensitivity and specificity of the Chinese version of the MoCA is limited due to a number of methodological issues. Kandiah *et al.* (2014) was rated as low quality due to a failure to report the time interval between the reference standard and the index test, the small sample size included in the study (n = 95), and the non-blinding of assessors, all of which could have impacted on estimates of diagnostic accuracy. Although some of these issues were relevant to Chen *et al.* (2013), this study included a larger sample size (n = 616), and reported that the reference standard and index test were completed on the same day which reduces biased estimates of test performance.

Portuguese (Brazilian)

The diagnostic accuracy of the Portuguese (Brazilian) version of the MoCA for detecting MCI and dementia in Parkinson's disease was evaluated in three studies. The findings from the first study (Almeida *et al.,* 2019) suggested 85% sensitivity and 72% specificity at a cut off of 17.5 for detecting dementia, and 82% sensitivity and 76% specificity at a cut off of 22.5 for detecting MCI, indicating that the MoCA has fairly good ability to screen for both

MCI and dementia in patients with Parkinson's disease. The findings from the second study (Sobreira *et al.*, 2015) indicated better accuracy in detecting dementia among those who had Parkinson's disease, with 94% sensitivity and 68% specificity at a cut-off of 21, however demonstrated less accuracy in screening for MCI with 84% sensitivity and 27% specificity at a cut-off of 26. In the third study (Camargo *et al.*, 2016), a cut-off score of 19 provided the best balance between sensitivity (88%) and specificity (89%) for detecting dementia in patients with Parkinson's disease. The authors suggested that as the MoCA is used as a screening test, a cut-off score with a higher sensitivity is perhaps more useful. Similar to Sobreira *et al.* (2015), they found that a cut-off score of 21 provided higher sensitivity (93%) however specificity was low (56%). All three studies were rated as low quality due to the methodological limitations discussed above in addition to small sample size and the non-blinding of assessors which were common to all three studies. Furthermore, although sensitivity was generally reported as good, specificities were relatively low, compromising diagnostic accuracy.

English

A total of five studies investigated the diagnostic accuracy of the English version of the MoCA. Two of the studies examined the diagnostic accuracy of the MoCA in detecting dementia in Parkinson's disease. Both studies (Dalrymple-Alford *et al*., 2010; Hoops *et al*., 2009) suggested that the MoCA has good overall discriminant validity as a screening tool for the detection of dementia. It is notable however that different optimal cut-off scores were found in these studies, with a higher cut-off score of 24.5 (sensitivity = 82%, specificity = 75%) reported in the study by Hoops *et al.* (2009) compared to a cut-off of 21 (sensitivity = 81%, specificity = 95%) in Dalrymple-Alford *et al*. (2010). This variability may be explained by differences in demographic characteristics within the studies. For example, the mean education (>16 years) reported by Hoops *et al.* (2009) was considerably higher than the mean education (>12 years) reported in the study by Dalrymple-Alford *et al*. (2010). The mean age in the Hoops *et al.* (2009) study was also around five years lower than that reported in the study by Dalrymple-Alford *et al*. (2010). The population in the study by Hoops *et al.* (2009) may not be representative of the majority of patients with Parkinson's disease and this may affect the generalisability of the results.

Three of the studies examined the diagnostic accuracy of the English version of the MoCA in detecting MCI in Parkinson's disease. The findings from the first study (Hendershott *et al.*, 2019) suggested 80% sensitivity and 68% specificity at a cut off of 26, indicating that

the MoCA is a sensitive screening tool for detecting MCI in Parkinson's disease. Marras *et al.* (2013) also reported a cut-off score of 26 with a sensitivity of 83% and a relatively low specificity of 44%. The authors reported that a desirable combined sensitivity and specificity could not be achieved. In the third study (Brown *et al.*, 2016), the optimal cut-off score was slightly lower than the others at 25, with a sensitivity of 77%, and a specificity of 79%. The studies by Hendershott *et al.* (2019) and Brown *et al.* (2016) were rated as low quality due to the methodological issues discussed above, which compromise the validity of the conclusions that can be made as to diagnostic accuracy. Marras *et al.* (2013), on the other hand, reported the time interval between the reference standard and the index test, and ensured the blinding of assessors when administering the MoCA, which makes the conclusions drawn from this study more robust. It was, therefore, given a medium quality rating.

Discussion

The aim of the current review was to investigate the diagnostic accuracy of English and non-English versions of the MoCA in diagnosing MCI and/or dementia in patients with PD. Although studies generally reported good sensitivity for detecting MCI and dementia across different language versions of the MoCA, specificities in many of the studies were fairly low, reducing diagnostic accuracy. Study findings were also compromised by a number of methodological flaws. Indeed, across most of the studies reviewed there was a lack of information on: blinding of assessors, the time period between administration of the MoCA and the reference standard, a priori defined cut-off thresholds, and power calculation and intended sample size. As a consequence, it is not possible to know the potential for biased estimates of test performance and thus diagnostic accuracy. Future studies investigating the MoCA would benefit from including this information.

The review also found that optimal cut-off thresholds across studies, even within the same language versions, varied. Variance in optimal cut-off scores between studies may have been due to differences in populations across studies, such as age, education, and cultural factors (Lewis *et al*., 2009; O'Driscoll & Shaikh, 2017). Indeed, the mean age across studies differs by over 16 years between the youngest and oldest sample. Large variances also exist in mean education across studies ranging from 5.5 to 16.2 years. In addition, the proportion of patients with cognitive impairment in the samples varied widely across studies from 22% (Sobreira *et al.*, 2015) to 82% (Camargo *et al.*, 2016) which may also have impacted on the differences in findings. To be fit for clinical use, screening tools must have statistically

robust cut-off scores to allow distinction between the presence or absence of cognitive impairment (Lezak *et al*., 2004). The variance in obtained optimal cut-off scores found in this review is an issue potentially compromising the clinical utility of the MoCA in PD populations. This could be addressed in future research through the use of a priori defined cut-off thresholds (Leeflang *et al*., 2008).

In several of the studies, particularly the non-English language studies, the optimal diagnostic cut-off scores for detecting MCI or dementia in PD were lower than previously published normative data for the MoCA (Nasreddine *et al.,* 2005). Differences in demographics in the original study when compared to the studies included in this review may account for this variance. The original validation data was based on English and French versions of the MoCA and although the mean age in the original study was higher than most of the studies included in this review, the mean years of education was also significantly higher. The MoCA does not account for premorbid functioning or years of education which limits the ability of clinicians to draw conclusions about whether an individual's cognitive decline is greater than expected for their age and years of education (Gagnon *et al*., 2013). Given that some items in the MoCA are likely to be influenced by education (i.e. phonemic verbal fluency, verbal abstraction), it follows that individuals with lower levels of education are likely to score lower on the MoCA (Nasreddine *et al.,* 2005). Furthermore, years of education may not be cross-culturally equivalent, with many individuals in middle Eastern and Asian countries having received somewhat less education than those in Western countries (Lee *et al*., 2008). Premorbid functioning and culture have been shown to have an impact on other screening tool performance (Pedrazaa *et al.,* 2012), and should be measured in relation to the MoCA.

Similar findings were reported in a review study investigating the cross-cultural applicability of the MoCA in screening for MCI (O'Driscoll & Shaikh, 2017). A wide range of cut-off scores were found across and within countries, and as a consequence individuals could be considered as cognitively impaired or within the 'normal range' using the same cut-off score. Thus, the criteria to diagnose cognitive impairment may differ between countries based on educational and cultural variation, and the use of a universal cut-off score may lead to a sampling bias and inaccurate conclusions about an individual's cognitive ability (O'Driscoll & Shaikh, 2017). This highlights the need for population-based norms for each language version of the MoCA, and the need for further research on population characteristics other than age, education, and culture which may influence MoCA scores.

Although not specifically highlighted within the studies, it should be noted that level of motor impairment may also have an influence on test performance (Koski, 2013). Functional limitations that are common to PD populations may impact the response process and impede assessment of cognitive ability. Interestingly, Nazem *et al.* (2009) found that severity of motor impairment in PD populations was not associated with performance on the MoCA, particularly those items requiring motor skills. This would suggest that PD symptoms do not preclude assessment of cognitive ability using the MoCA. Ultimately, decisions about a patient's ability to complete a valid cognitive assessment will be based on the clinician's clinical judgment with some clinicians choosing to omit problematic items and interpret the resulting score with caution (Koski, 2013).

There are a wide range of cognitive screening tests available for detecting MCI and dementia however several diagnostic accuracy studies have shown that the MoCA performs favourably (Roalf *et al*., 2013; Tsio *et al*., 2015). In PD populations, research suggests that when compared with the Mini Mental State Examination (MMSE), which is amongst one of the most widely used screening tests, the MoCA is more sensitive and accurate, particularly in relation to detecting early cognitive impairment. The ability of the MoCA to detect a decline in the cognitive functions commonly affected in PD (i.e. reasoning, planning, and executive functions) has been widely reported as the main advantage of the MoCA over the MMSE (Camargo *et al.*, 2016; Chou *et al*., 2010; Dalrymple-Alford *et al*., 2010; Hoops *et al*., 2009). Indeed, the use of the MMSE as a screening test in PD has been challenged due to its limited executive function assessment (Bugalho & Vale, 2011).

These concerns have led to some doubts about the applicability of the MDS-TF criteria for diagnosing MCI and dementia in Parkinson's disease since the cognitive assessment used to determine the presence of cognitive impairment is based on the MMSE (Ohta *et al.,* 2014). Although it is widely acknowledged that the development of specific guidelines for diagnosing MCI and dementia in PD populations is an important milestone, it is essential that cognitive impairment in this population is detected early and diagnosed accurately. Since the development of the MDS-TF criteria, many studies have assessed its validity and have reported variable findings with most concluding that the accuracy of the MMSE was not satisfactory and that it lacked sensitivity (Barton *et al.,*2012; Isella *et al.,* 2014; Ohta *et al.,* 2014). Despite this, some studies have found that the MDS-TF criteria are more sensitive for diagnosing cognitive impairment than the previously widely used Diagnostic and

Statistical Manual of Mental Disorders, 4th edition (DSM-IV) definition (Martinez-Martin *et al.,* 2011). In any case, research suggests that the current MDS-TF criteria for diagnosing cognitive impairment in PD populations requires refinement.

Limitations of the Review

Due to the strict inclusion criteria, only a small number of studies were included in this review. Although a comprehensive and rigorous search strategy was implemented, there is always a possibility that papers may have been missed. The absence of a second rater for study selection is another clear limitation. In addition, the review included only English language studies which may have led to the exclusion of relevant research. In relation to the quality assessment process, the overall quality ratings (i.e. high, medium, or low) given to the studies was to assist the reader's interpretation of study quality however the cut-off scores used to categorise study robustness were arbitrary. Moreover, although the use of total scores can help to summarise the quality of each of the papers, reliance on total scores may fail to identify the studies which are at an increased risk of bias in some key areas and thus analysis of individual components of methodological quality should be considered (O'Conner et al., 2015).

Areas for Future Research and Clinical Practice

The variability of cut-off scores found in this review both across and within different language versions of the MoCA for detecting MCI and dementia in PD populations highlights the need for caution in applying them without first taking into consideration the effects of age, education, and cultural difference (Rossetti *et al.,* 2011). The low quality of diagnostic accuracy studies emphasises the importance of using the MoCA as a screening tool only, and interpreting test scores within the context of a more comprehensive assessment (Wong *et al.,* 2015). There is a clear need to build up an evidence base for specific cut-off scores for specific language versions of the MoCA by seeking to evaluate the performance of cut-off thresholds reported in previous studies rather than simply reporting the cut-off score which maximises test performance in any given sample. Improving the validity of diagnostic accuracy studies through adherence to STARD guidelines is also essential for future research.

Conclusion

Although the studies included in this review generally reported good sensitivity for detecting MCI and dementia across different language versions of the MoCA, specificities in many of the studies were fairly low, particularly for MCI, reducing diagnostic accuracy. Moreover, optimal cut-off thresholds, even within particular language versions, varied, compromising the clinical utility of the MoCA in PD populations. Future research should address the methodological limitations highlighted by this review through adherence to STARD guidelines.

References

Aarsland. D., Andersen, K., Larsen, J.P., Lolk, A., Kragh-Sorensen, P. (2003). Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Arch Neurol, 60*, 387–392.

Almeida, K. J., Carvalho, L., Monteiro, T., Gonçalves, P. & Campos-Sousa, R. N. (2019). Cut-off points of the Portuguese version of the Montreal Cognitive Assessment for cognitive evaluation in Parkinson's disease. *Dementia & neuropsychologia*, *13* (2), 210–215.

Alzheimer's Society (2018a). What is dementia? Available from: https://www.alzheimers.org.uk/about-dementia/types-dementia/what-dementia

Alzheimer's Society (2018b). Mild Cognitive Impairment (MCI). Available from: https://www.alzheimers.org.uk/about-dementia/types-dementia/mild-cognitive-impairment-mci

Barton, B., Grabli, D., Bernard, B., *et al.* (2012). Clinical validation of movement disorder society–recommended diagnostic criteria for Parkinson's disease with dementia. *Mov. Disord., 27*, 248-253.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., *et al*. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *British Medical Journal, 351*, 1-9.

Brown, D. S., Bernstein, I. H., McClintock, S. M., Munro Cullum, C., Dewey, R. B., Jr, Husain, M. & Lacritz, L. H. (2016). Use of the Montreal Cognitive Assessment and Alzheimer's Disease-8 as cognitive screening measures in Parkinson's disease. *International journal of geriatric psychiatry*, *31*(3), 264–272.

Bugalho, P. & Vale, J. (2011). Brief cognitive assessment in the early stages of Parkinson disease. *Cognitive and Behavioral Neurology, 24* (4), 69–173.

Camargo, C., Tolentino, E. S., Bronzini, A., Ladeira, M. A., Lima, R., Schultz-Pereira, G. L. & Young-Blood, M. R. (2016). Comparison of the use of screening tools for evaluating

cognitive impairment in patients with Parkinson's disease. *Dementia & neuropsychologia*, *10* (4), 344–350.

Chen, L., Yu, C., Fu, X., Liu, W., Hua, P., Zhang, N. & Kuo, S. (2013). Using the Montreal Cognitive Assessment Scale to screen for dementia in Chinese patients with Parkinson's Disease. *Shanghai archives of psychiatry*, *25* (5), 296–305.

Chou, K.L., Amick, M.M., Brandt, J., *et al*. (2010) A recommended scale for cognitive screening in clinical trials of Parkinson's disease. Mov Disord. 25(15), 2501-7.

Cullen, B., O'Neill, B., Evans, J.J., Coen, R.F. & Lawlor, B.A. (2007). A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry, 78*, 790-799.

Dalrymple-Alford, J. C., MacAskill, M. R., Nakas, C. T., *et al.* (2010). The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology*, *75* (19):1717–1725.

De Lau, L.M. & Breteler, M.M. (2006). Epidemiology of Parkinson's Disease. *The Lancet Neurology, 5* (6), 525–535.

Emre, M., Aarsland, D., Brown, R., *et al.* (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord, 22*(12),1689-707.

Federico, A., Maier, A., Vianello, G., *et al.* (2015). Screening for Mild Cognitive Impairment in Parkinson's Disease: Comparison of the Italian Versions of Three Neuropsychological Tests. *Parkinson's disease*, *2015*, 681976.

Florkowski, C. (2008). Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. The Clinical biochemist. *Reviews / Australian Association of Clinical Biochemists, 29* (1), 83-7.

Gagnon, G., Hansen, K., Woolmore-Goodwin, S., *et al*. (2013). Correcting the MoCA for Education: Effect on Sensitivity. The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques. 40. 678-83.

Gifford, D. R. & Cummings, J.L. (1999). Evaluating Dementia Screening Tests: Methodologic standards to rate their performance. *Neurology, 52*(2), 224-227.

Hendershott, T.R., Zhu, D., Llanes, S., Zabetian, C.P., Quinn, J., Edwards, K.L., Leverenz, J.B., Montine, T., Cholerton, B. & Poston, K.L. (2019). Comparative sensitivity of the MoCA and Mattis Dementia Rating Scale-2 in Parkinson's disease. *Mov Disord., 34*, 285-291.

Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B. & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, *73* (21), 1738–1745.

Isella, V., Mapelli, C. Siri, C., *et al*. (2014). Validation and attempts of revision of the MDS-recommended tests for the screening of Parkinson's disease dementia, *Park. Relat. Disord. 20* (1), 32-36.

Kandiah, N., Zhang, A., Cenina, A. R., Au, W. L., Nadkarni, N. & Tan L. C. (2014). Montreal Cognitive Assessment for the screening and prediction of cognitive decline in early Parkinson's disease. *Parkinsonism and Related Disorders*, *20* (11), 1145–1148.

Knottnerus, J.A. & Muris, J.W. (2003). Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol*, *56*, 1118–28.

Koski, L. (2013). Validity and Applications of the Montreal Cognitive Assessment for the Assessment of Vascular Cognitive Impairment. *Cerebrovasc Dis, 36*, 6-18.

Larner, A. (2012). Screening utility of the Montreal Cognitive Assessment (MoCA): In place of – or as well as – the MMSE? *International Psychogeriatrics, 24* (3), 391-396.

Lee, J.Y., Lee, D.W., Cho, S.J., *et al*. (2008). Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. *J Geriatr Psychiatry Neurol, 21*, 104–110.

Leeflang, M.M., Moons, K.G., Reitsma, J.B., *et al*. (2008). Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clin Chem*, *54*, 729–37.

Levy, G., Tang, M.X., Louis, E.D., *et al*. (2002). The association of incident dementia with mortality in PD. *Neurology, 59*, 1708–1713.

Lewis, C., Koyasu, M., Oh, S., Ogawa, A., Short, B. & Huang, Z. (2009) Culture, executive function, and social understanding. *New Dir Child Adolesc Dev*, *123*, 69-85.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J. & Fischer, J. S. (2004). Neuropsychological assessment (4th ed.). New York, NY, US: Oxford University Press.

Litvan, I., Aarsland, D., Adler, C.H., *et al.* (2011). MDS task force on mild cognitive impairment in Parkinson's disease: critical review of PDMCI. *Mov Disord, 26*, 1814–1824

Litvan, I., Goldman, J. G., Tröster, A. I., *et al*. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Movement disorders: official journal of the Movement Disorder Society*, *27* (3), 349–356.

Marras, C., Armstrong, M. J., Meaney, C. A., *et al.* (2013). Measuring mild cognitive impairment in patients with Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*, *28* (5), 626–633.

Martinez-Martin, P., Falup-Pecurariu, C., Rodriguez-Blazquez, C., *et al.* (2011). Dementia associated with Parkinson's disease: applying the Movement Disorder Society Task Force criteria. *Parkinsonism Relat Disord, 17*, 621–624.

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., *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.

National Institute for Health and Care Excellence (NICE). (2006). Dementia: Supporting people with dementia and their carers in health and social care. Social Care Institute for Excellence. Available from: https://www.nice.org.uk/Guidance/CG42

National Institute for Health and Care Excellence (NICE). (2014). Developing NICE guidelines: The manual 2014. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview

Nazem, S., Siderowf, A. D., Duda, J. E., *et al.* (2009). Montreal cognitive assessment performance in patients with Parkinson's disease with "normal" global cognition according to mini-mental state examination score. *Journal of the American Geriatrics Society*, *57* (2), 304–308.

O'Connor, S.R., Tully, M.A., Ryan, B. *et al.* (2015). Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes 8,* 224.

O'Driscoll, C. & Shaikh, M. (2017). Cross-Cultural Applicability of the Montreal Cognitive Assessment (MoCA): A Systematic Review. *Journal of Alzheimer's disease, 58* (3), 789-801.

Ohta, K. Takahashi, K., Gotoh, J., *et al.* (2014). Screening for impaired cognitive domains in a large Parkinson's disease population and its application to the diagnostic procedure for Parkin- son's disease dementia, Dement. *Geriatr. Cogn. Disord. Extra 4* (2), 147-159.

Ozdilek, B. & Kenangil, G. (2014). Validation of the Turkish Version of the Montreal Cognitive Assessment Scale (MoCA-TR) in Patients With Parkinson's Disease, *The Clinical Neuropsychologist*, *28* (2), 333-343.

Pedrazaa, O., Clark, J. H., O'Bryant, S. E., *et al.* (2012). Diagnostic validity of age and education corrections for the mini-mental state examination in older African Americans. *The International Neuropsychological Society, 60*, 328–331.

Roalf, D. R., Moberg, P. J., Xie, S. X., Wolk, D. A., Moelter, S. T., & Arnold, S. E. (2013). Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *9* (5), 529–537.

Shulman, J. M., De Jager, P. L., & Feany, M. B. (2011). Parkinson's disease: genetics and pathogenesis. *Annual Review of Pathology: Mechanisms of Disease, 6*(1), 193–222.

Sobreira, E., Pena-Pereira, M.A., Eckeli, A.L., Sobreira-Neto, M.A., Chagas, M.H., Foss, M.P., *et al.* (2015). Screening of cognitive impairment in patients with Parkinson's disease: diagnostic validity of the Brazilian versions of the Montreal Cognitive Assessment and the Addenbrooke's Cognitive Examination- Revised. *Arquivos de Neuro-Psiquiatria, 73* (11), 929-933.

Tsoi, K.K.F., Chan, J.Y.C., Hirai, H.W., *et al.* (2015). Cognitive Tests to Detect Dementia: A Systematic Review and Meta-analysis. *JAMA Intern Med, 175*(9), 1450–1458.

Whiting, P., Rutjes, A.W., Reitsma, J.B., *et al.* (2004) Sources of Variation and Bias in Studies of Diagnostic Accuracy: A Systematic Review. *Ann Intern Med, 140*, 189–202.

Wong, A., Law, L., Liu, W., *et al.*, (2015). Montreal Cognitive Assessment: One cut-off never fits all. *Stroke, 46*, 3547-3550.

CHAPTER 2: Major Research Project

An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

Hollie Thomson[1]*

[1]Mental Health and Wellbeing, Institute of Health and Wellbeing, University *of* Glasgow

*Address for Correspondence:
University *of* Glasgow
Mental Health and Wellbeing
Gartnavel Royal Hospital
1st Floor, Administration Building
1055 Great Western Road
Glasgow
G12 0XH
Email: Hollie.Thomson@ggc.scot.nhs.uk

Prepared in accordance with requirements for submission to the International Journal of Geriatric Psychiatry (Appendix 1.1)

Submitted in part fulfilment of the requirements for the Degree of Doctorate in Clinical Psychology (D.Clin.Psy)

Word Count: 6,976

Abstract

Background: Dementia is a leading cause of disability amongst older adults in the developed world. In recent years, an increasing emphasis has been placed on the early detection and diagnosis of dementia, and timely access to post-diagnostic support. Cognitive screening tests are essential tools in facilitating the process of early detection and dementia diagnosis and are currently widely used in clinical practice. The Addenbrooke's Cognitive Examination-III (ACE-III) is one such tool. Despite it being recommended in several evidence-based guidelines and being widely used in the NHS, the rater reliability of the ACE-III has never been formally evaluated. **Aims:** To investigate rater accuracy in scoring the ACE-III both in terms of its total and sub-category scores across different raters and by the same raters at two different time points. A secondary exploratory analysis examines whether scoring accuracy is affected by participants' experience of using the ACE-III and/or whether they have had formal training on how to administer and score the ACE-III. **Methods:** A filmed vignette of the ACE-III being administered to an older adult actor (mock patient) was used to assess scoring accuracy across different raters. The vignette has pre-determined 'true' scores. Participants were asked to view the filmed vignette whilst simultaneously completing an ACE-III scoring sheet. Following a two-month period, participants were invited back to view and score the same vignette again to assess *intra*-rater reliability. Participants were NHS staff working in Older People's Community Mental Health Teams who routinely administer and score the ACE-III as part of their clinical practice. **Results and Conclusions:** The *inter*- and *intra*-rater scoring accuracy of the ACE-III is generally good, with error mainly accounted for by the domains which require more subjective judgements, namely, the visuospatial and language tests. Health professionals should therefore take these findings into account when scoring the ACE-III, and utilise the ACE-III administration and scoring guide to help improve accuracy. If all health professionals score the ACE-III in a consistent manner, the accurate and early identification and management of those individuals with a dementia will be improved.

Introduction

Background

Dementia is an umbrella term for a range of progressive conditions that are characterised by global impairment of cognitive functioning (Alzheimer's Society, 2018). Dementia is a leading cause of disability amongst older adults in the developed world, and constitutes one of the most significant, and costly, challenges currently facing health and social care services (Milne *et al* ., 2008). In the UK alone, there are approximately 850,000 people currently living with a dementia (Alzheimer's Society, 2018). Given that age is the biggest risk factor for developing a dementia, and that the average life expectancy is increasing, it is not surprising that the incidence and prevalence of dementia is expected to rise rapidly over the next several decades. Indeed, it is estimated that by 2051 there will be around two million people in the UK living with a dementia (Alzheimer's Society, 2018). Over the past decade, the Scottish Government has made dementia a public health priority and has set out a range of commitments and objectives to improve the quality of care for individuals with a dementia, with increasing emphasis being placed on early detection and diagnosis, and timely access to post-diagnostic support (Scottish Government, 2017). Early detection of dementia is essential to ensuring early intervention and management of the condition, and improving quality of life (Scottish Government, 2017).

The Role of Cognitive Screening Tests in Dementia Diagnosis

The National Institute for Health and Care Excellence (NICE, 2018) recently published guidelines on the diagnosis and management of dementia which stipulate the use of a validated brief structured cognitive screening test as part of the initial stage of a comprehensive dementia assessment. Indeed, the assessment of cognitive functioning is central to the process of early detection and diagnosis of dementia and is therefore arguably one of the most important assessments made by clinicians in clinical practice (Machado *et al* ., 2015). In the context of time constraints, the efficient early detection of dementia requires the use of an objective cognitive screening test which is brief, reliable, easy to administer, and acceptable to patients (Cullen *et al*., 2007; Villarejo & Puerta-Martin, 2011). It is not surprising then that in recent years there has been an increase in the number of dementia screening tests proposed and studied.

The main aim of a cognitive screening test is to provide information about the presence, or absence, of a cognitive impairment (Lezak *et al*., 2004). This is often inferred from the

individuals' score on the test compared to referenced norms (Cullen *et al*., 2007). A central feature of most screening tests is the 'cut off' score, that is, the point at which an individual's score moves from being within the 'normal' range to being within the 'impaired' range, i.e. indicating the presence of a dementia. The success of a screening test in achieving this aim undoubtedly lies in its psychometric robustness (Cullen *et al*., 2007). Cognitive screening tests should have good sensitivity and specificity to ensure that individuals with a cognitive impairment are not missed and that individuals without a cognitive impairment are not misidentified as impaired and referred on for further testing unnecessarily (Stolwyk *et al*., 2014).

One of the most important attributes of any assessment instrument is its reliability, that is, the extent to which the test produces consistent measurements when repeated under replicate conditions (Van Belle & Arnold, 2000). There are a number of factors which may influence the reliability of a test including individual factors (i.e. mood, fatigue, and motivation), environmental factors (i.e. temperature and noise), and rater factors (i.e. training, experience, judgement). These factors are sources of measurement error in the assessment process. If these measurement errors did not exist, one would expect that an individual would obtain the same test score, their 'true' score, each time they took the test. It is generally accepted however that the observed score on a test equates to the 'true' score plus some degree of measurement error (Trochim, 2006). Thus, the extent to which a cognitive test can minimise the impact of measurement errors, and ensure that the score obtained is as close to the individuals 'true' score as possible, is an indication of its reliability (Trochim, 2006). *Inter*-rater reliability is the level of scoring consistency between raters evaluating the same test. However, agreement between raters does not necessarily equate to scoring accuracy when evaluating a test, since raters may show high levels of agreement but show low levels of accuracy. Therefore, rater accuracy in scoring tests across different raters is an important aspect of reliability that must be taken into consideration in *inter*-rater reliability studies. *Intra*-rater reliability assesses scoring consistency by the same raters at two different time points.

The ACE-III as a Cognitive Screening Test for Dementia

The Addenbrooke's Cognitive Examination, now in its third edition (ACE-III), is one of the cognitive screening tests recommended by the Scottish Intercollegiate Guideline Network (SIGN) in its guideline on the Management of Patients with Dementia (SIGN, 2006). The ACE-III is comprised of five subscales, each representing a cognitive domain; Attention,

Memory, Fluency, Language, and Visuo-spatial. A maximum score of 100 can be obtained, with higher scores indicative of better memory and cognitive performance. The ACE-III is designed to be sensitive to the early stages of dementia. Cognitive domains within the ACE-III have been validated against a battery of standardized neuropsychological tests, with high levels of correlation between the domain scores and targeted tests used in the assessment of attention, language, verbal memory and visuospatial function (Hsieh *et al*., 2013). The ACE-III also compared favourably with its predecessor, the ACE-R, with similar levels of sensitivity and specificity, in the assessment of cognitive deficits in people with Alzheimer's disease and fronto-temporal dementia (Hsieh *et al*., 2013). The ACE-III shows high sensitivity and specificity for mild cognitive impairment (MCI) at a cut-off of 88 (sensitivity = 1.0; specificity = 0.96) and dementia at a cut-off of 82 (sensitivity = 0.93; specificity = 1.0) (Hsieh *et al*., 2013). It takes approximately 15 minutes to administer the ACE-III and it therefore meets the requirements of a screening test which is time efficient (Cullen *et al*., 2007). There are three versions of the ACE-III available to reduce any practice effects with repeat testing. Appropriate use of the ACE-III requires clinicians who are able to accurately administer, score, and interpret the test. The ACE-III was devised for use by a range of professional groups who work with patient populations that require cognitive screening including General Practitioners (GPs), Psychiatrists, Nurses, Occupational Therapists (OTs), and Clinical Psychologists (CPs). Despite this, validation studies for the ACE-III have generally utilised Psychologists, Neurologists, and Neuropsychologists to administer and score the test (Takenoshita *et al.,* 2019).

The Current Study

Although the psychometric properties of cognitive screening tests are widely reported, there is a dearth of research exploring the rater reliability and scoring accuracy of such tests among clinicians who routinely use them in clinical practice (Newman *et al* ., 2018). Previous studies examining test scoring on cognitive screening measures have found high rates of error (Crawford, 2010; Kozora, 2018; Sullivan, 2000) which highlights the importance of usability research, and ensuring that tests are used appropriately in real-world settings. Similarly, the consistency with which a test is scored by a range of clinicians is rarely reported. This is particularly relevant when there are items on a test which require more subjective scoring judgements.

Despite being recommended in several evidence-based guidelines as a screening test for dementia, and being widely used in the NHS, the rater reliability of the ACE-III has never

been formally evaluated. As with any method of objective testing, it is essential to establish whether the ACE-III is accurately and consistently scored across different raters and by the same raters at different time points. Indeed, this is highly relevant to ensuring the accurate and early identification and management of those individuals with a dementia.

Aims

This study investigated rater accuracy in scoring the ACE-III. It explored how accurately the ACE-III was scored both in terms of its total and sub-category scores across different raters and by the same raters at two different time points. It also considered whether scoring accuracy was affected by participants' experience of using the ACE-III and/or whether they had formal training on how to administer and score the ACE-III.

Research Questions

The specific research questions in relation to this study are:

(i) Do participants' ACE III total scores differ significantly from the ACE III 'true' score?

(ii) Do participants' ACE-III sub-category scores differ significantly from the ACE-III 'true' scores?

(iii) What is the test-retest scoring consistency across two different time points?

Secondary Exploratory Analysis

A secondary exploratory analysis was conducted to examine whether scoring accuracy was affected by participants' experience of using the ACE-III and/or whether they had received formal training on how to administer and score the ACE-III.

Methods

Participants

The participants in this study were NHS staff working in Older People's Community Mental Health Teams in Greater Glasgow and Clyde who routinely administer and score the ACE-III as part of their clinical practice. Participants were recruited from several groups of professionals including, Clinical Psychologists, Community Psychiatric Nurses (CPNs), Occupational Therapists (OTs), and Care Home Liaison Nurses (CHLNs).

Eligibility Criteria

To be considered eligible for inclusion in this study, participants were required to have administered and scored the ACE-III independently in clinical practice on at least one occasion with an older adult patient. Participants were excluded if they had not administered the ACE-III in clinical practice. Some participants had previously received training on how to administer and score the ACE-III; this did not preclude them from inclusion in the study.

Justification of Sample Size

A-priori sample size calculation was completed using G*Power (Faul *et al*., 2007) with an assumed power of 0.8 and an error value of 0.05. Calculations were based on the method of data analysis required to answer the main research question; one sample *t*-tests (or the non-parametric equivalent) to compare differences between participants' ACE-III total and domain scores and the predetermined 'true' scores. G*Power (Faul *et al*., 2007) indicated that a minimum of 34 participants would be required to detect a medium effect size. A medium effect size was deemed to be sufficient enough to detect a clinically significant difference, that is, a difference large enough to alter interpretation of the ACE-III results (Crawford, 2010).

Measures

The Addenbrooke's Cognitive Examination-III (ACE-III) was the primary outcome measure.

Design Procedure

Permission was given to use a filmed vignette from the online NHS Education for Scotland (NES) ACE-III Trainer programme for the purposes of this study. The University of Glasgow media services originally filmed and produced the vignette. The vignette is of a Clinical Psychologist administering the ACE-III to an older adult actor (mock patient). The vignette has pre-determined 'true' scores which had been agreed upon by two experienced Clinical Psychologists. Given that the scripts for the vignettes were written jointly by Clinical Psychologists with extensive neuropsychological experience within older adult services, and that the answers were also co-produced, there was 100% agreement on what the 'true' scores should be. Pre-determined 'true' scores allow for scoring consistency to be investigated separately from administration consistency. The vignette has a 'true' total score of 63/100.

Research Procedure

Participants were recruited from Older People's Community Mental Health Teams across NHS Greater Glasgow and Clyde. Potential participants were invited to take part in the study via email and received an information sheet about the study (Appendix 2.1). Participants who opted in to the study were invited to a research session lasting approximately one hour, where they watched a vignette whilst simultaneously completing an ACE-III scoring sheet. The vignette was shown to participants in groups and played on a projector screen. Participants were asked to view the vignette once and were not be permitted to pause or rewind it, reflecting actual clinical practice. The ACE-III administration and scoring guide was made available to participants on request. Participants were not permitted to consult each other regarding scoring and were blinded from each other's ratings. Participants were also asked not to discuss their scoring after the session. In addition to scoring the vignettes, participants were also asked to complete an information form detailing their profession, ACE-III experience, and whether or not they had completed any formal training in using the ACE-III (Appendix 2.2). A standardised set of instructions were given verbally to each group of participants prior to commencing the study (Appendix 2.3). Participants were also given the correct orientation information for the vignette (Appendix 2.4). Consent forms were signed at the group session prior to participation (Appendix 2.5). To evaluate *intra*-rater reliability, participants were invited back after a period of approximately two months to view and score the same vignette again, under the same conditions and in an identical manner to the first time point. Participants were asked to confirm ongoing consent to participate in the study given the passage of time. Participants did not have access to their previous ratings. Participants were also asked for an update as to whether they had received any additional ACE-III training since the previous test session.

Ethical Approval

Prior to the commencement of the study, ethical approval was sought and obtained from the College of Medical, Veterinary and Life Sciences (MVLS) Ethics Committee (Appendix 2.6). Approval was sought and obtained from NHS Greater Glasgow and Clyde R&D (Appendix 2.7).

Data Analysis

Descriptive statistics including proportions and standard deviations were used to analyse the variation of test scores from the 'true' scores. Percentages were also used to determine how many participants correctly identified the 'true' total, domain, and sub-domain scores. The

number of participants scoring consistently with the true score, within two points, or deviating by three or more points is reported. A two point discrepancy is reported as this was considered likely to be acceptable in clinical practice and was also used by Crawford (2010) in a previous investigation of scoring reliability of the ACE-R.

All statistical analyses were carried out using IBM® SPSS Statistics, version 26.0. Initial investigations revealed that the data did not meet the assumptions of normality and a non-parametric approach to statistical analysis was taken. One sample Wilcoxon Signed-Rank tests were used to compare whether participants' ACE-III total scores differed significantly from the predetermined 'true' scores, and to investigate whether the results obtained differed significantly from the 'true' domain scores. Paired samples Wilcoxon Signed-Rank tests were used to compare differences between participants' total and domain scores between session one and session two. Correlational methods were used to explore *intra*-rater reliability, that is, whether participants' ACE-III scores were consistent across two different time points. Mann-Whitney U tests were used to explore whether scoring accuracy was affected by participants' experience of using the ACE-III and/or whether they had received formal training on how to administer and score it.

Results

Forty-one participants (38 females; 3 males) consented to take part in the study; 31 Community Psychiatric Nurses (CPN's), 6 Occupational Therapists (OT's), 1 Clinical Psychologist (CP), and 3 Care Home Liaison Nurses. Participants' length in post ranged from one year or less to 24 years (M = 8.6, SD = 6.9). All participants attended the first research session of the study. One participant did not sum items to provide domain and total scores and this was therefore treated as missing data and excluded from relevant analyses. Of the 41 participants who completed the first part of the study, 37 attended the second research session approximately two months later.

Scoring Accuracy for Session One

Table 1 summarises the accuracy figures for participant ACE-III total, domain, and sub-domain scores. Participant total scores for the ACE-III vignette ranged from 58 to 68 points (M = 63.05, SD = 2.3). Twenty percent of participant total scores matched the predetermined 'true' score (TS), while 48% of total scores deviated from the TS by 1-2 points, and 32% deviated from the TS by 3-5 points. No participant total scores deviated from the TS by

more than 5 points. Forty-five percent of participant ACE-III scores were higher than the TS, while around 35% were lower, indicating an overall tendency towards over-scoring. Scoring errors were highest within the language and visuospatial domains, with 63% and 76% of participant scores deviating from the TS respectively. At the sub-domain level, this deviation from the TS was mainly accounted for by scoring errors on the clock drawing item and on the single word repetition item, with 66% of participant scores deviating from the TS on both items. There was a tendency towards over-scoring within the language domain (M = 21.7, SD = .66) and on the clock drawing item (M = 1.4, SD = .58) while a tendency for under-scoring was observed within the visuospatial domain (M = 7.4, SD = .77) and on the single word repetition item (M = 1.7, SD = .52). Errors were present but observed less for the fluency (20%), memory (29%), and attention (46%) domains.

One sample Wilcoxon Signed-Rank tests were conducted to compare differences between participant total and domain ACE-III scores and the TS. The results indicated that there was a statistically significant difference between participant scores and the TS on the language (Z= -4.435, p < .001) and visuospatial (Z = 3.970, p <.001) domains, with large effect sizes. No statistically significant difference was found between participant scores and the TS on the attention, memory, or fluency domains. The results also indicated that there was not a statistically significant difference between participant total ACE-III scores and the TS. Table 2 presents the median, interquartile range, and test statistics.

Table 1 Accuracy for ACE-III total, domain, and sub-domain scores

	ACE-III True Score (TS)	Same as TS at Session 1 (S1)	+/- 2 points from TS (S1)	+/- 3 or more points from TS (S1)	Same as TS at Session 2 (S2)	+/- 2 points from TS (S2)	+/- 3 or more points from TS (S2)
ACE-III (%)							
Total	**63/100**	**20**	**48**	**32**	**24**	**65**	**11**
Attention (%)							
Orientation 1	2/5	100	-	-	100	-	-
Orientation 2	5/5	100	-	-	100	-	-
Registration	3/3	95	5	-	100	-	-
Numbers	3/5	71	29	-	71	29	-
Domain Total	**13/18**	**51**	**44**	**2**	**66**	**34**	**-**
Memory (%)							
Recall	3/3	100	-	-	100	-	-
Anterograde	6/7	98	2	-	100	-	-
Retrograde	3/4	90	10	-	95	5	-
Name & Address Recall	2/7	68	32	-	68	32	-
Name & Address Recog.	3/5	63	37	-	81	19	-
Domain Total	**17/26**	**71**	**24**	**5**	**67**	**30**	**3**
Fluency (%)							
Letters	2/7	93	7	-	95	5	-
Animals	2/7	90	10	-	84	16	-
Domain Total	**4/14**	**80**	**20**	**-**	**84**	**16**	**-**
Language (%)							
Comprehension Pencil/Paper	3/3	100	-	-	100	-	-
Comprehension Pictures	3/4	100	-	-	97	3	-
Object Naming	12/12	100	-	-	100	-	-
Reading	1/1	100	-	-	100	-	-
Proverb Repetition 1	1/1	100	-	-	100	-	-
Proverb Repetition 2	0/1	100	-	-	97	3	-
Single Word Repetition	1/2	34	66	-	32	62	6
Sentence Writing	0/2	98	2	-	97	3	-
Domain Total	**21/26**	**37**	**63**	**-**	**35**	**62**	**3**
Visuospatial (%)							
Dots	2/4	93	7	-	87	13	-
Letters	4/4	100	-	-	97	3	-
Clock	2/5	34	66	-	43	57	-
3D Cube	0/2	90	10	-	92	8	-
Infinity	0/1	100	-	-	100	-	-
				-			-
Domain Total	**8/16**	**24**	**76**	**-**	**30**	**70**	**-**

Table 2 Summary of one sample Wilcoxon Signed-Rank tests

	True Score (TS)	*Mdn*	*Interquartile Range*	Test statistics ($N = 40$) *Z*	*P*	*r*
Attention	13	13	12 - 13	-1.035	0.300	0.16
Memory	17	17.5	16.25 – 18	-1.078	0.281	0.17
Fluency	4	4	4 – 4	-0.378	0.705	0.06
Language	21	22	21 – 22	-4.435	0.000	0.70
Visuospatial	8	7	7 – 8	3.970	0.000	0.63
Total	63	63	62 - 64.75	-.189	0.850	0.03

Scoring Accuracy for Session Two

Participant total scores for the ACE-III vignette ranged from 60 to 66 points ($M = 63.19$, $SD = 1.6$). Twenty-four percent of participant total scores matched the predetermined TS, while 65% of total scores deviated from the TS by 1-2 points, and 11% deviated from the TS by 3 points. No participants deviated from the TS by more than 3 points. Forty-six percent of participant ACE-III scores were higher than the TS, while around 30% were lower, again indicating an overall tendency towards over-scoring.

Scoring errors were again highest within the language and visuospatial domains, with 65% and 70% of participant scores deviating from the TS respectively. At the sub-domain level, this deviation from the TS was again mainly accounted for by scoring errors on the clock drawing item and on the single word repetition item, with 68% and 57% of participant scores deviating from the TS respectively. There was a tendency towards over-scoring within the language domain ($M = 21.6$, $SD = .86$) and on the clock drawing item ($M = 1.4$, $SD = .50$) while a tendency for under-scoring was observed within the visuospatial domain ($M = 7.5$, $SD = .87$) and on the single word repetition item ($M = 1.7$, $SD = .56$). Errors were present but observed less for the fluency (16%), memory (33%), and attention (34%) domains. Table 1 summarises the percentage accuracy for participant ACE-III total, domain, and sub-domain scores.

Paired samples Wilcoxon Signed-Rank tests were conducted to compare differences between participant total and domain ACE-III scores at session one and session two. Table

3 presents the median, interquartile range, and test statistics. The results indicated that there was not a statistically significant difference between participant total ACE-III scores at session one and session two. There was a statistically significant different between participant scores on the fluency domain (Z = -2.309, p =.021) with a medium effect size. Small effect sizes that were not statistically significant were also observed for the memory, attention, language, and visuospatial domains.

Table 3 Differences in participant total and domain ACE-III scores between session one and session two

					Test statistics (N =37)		
	Mdn Session 1	*Interquartile Range*	*Mdn* Session 2	*Interquartile Range*	*Z*	*P*	*r*
Attention	13	12 - 13	13	13 -13	-0.250	0.803	0.04
Memory	17.5	16.25 - 18	17.5	17 - 18	-0.783	0.434	0.13
Fluency	4	4 -4	4	4 - 4	-2.309	0.021	0.37
Language	22	21 - 22	22	21 - 22	-0.361	0.718	0.06
Visuospatial	7	7 - 8	7	7 – 8	-0.968	0.333	0.16
Total	63	62 – 64.75	63	62 - 64	-0.350	0.726	0.06

Test-Retest Scoring Consistency

The data indicates that 46% of participants were consistent in their scoring of the ACE-III total between session one and session two. Only 13% of these participant scores matched the TS. To verify the level of agreement between *intra*-rater ACE-III total and domain scores, the intraclass correlation coefficient (ICC) was used. The ICC values were classified as follows: <0.4: poor reliability; 0.4-0.75: good reliability; >0.75: excellent reliability (Fleiss, Levin, & Paik, 2003). A good level of *intra*-rater reliability was found for ACE-III total scores - the ICC was .70 with a 95% confidence interval from .41 to .85 (F(35, 35) = 3.300, p < .001. Table 4 presents the ICC values for *intra*-rater total and domain scores.

Table 4 Reliability of total and domain scores: intraclass correlation coefficient (ICC)

	ICC (95% CI)
Attention	.06 (-.91, .53)
Memory	.49 (-.01, .74)
Fluency	.29 (-.28, .62)
Language	.56 (.13, .78)
Visuospatial	.58 (.18, .79)
Total	.70 (.41, .85)

Participant ACE-III Experience and Training

A secondary exploratory analysis was conducted to examine whether scoring accuracy was affected by participants' experience of using the ACE-III and/or whether they received training on how to administer and score it. The majority of participants (66%) reported having used the ACE-III in clinical practice for 5 years or more, and 88% reported using it at least once a month. Almost all of the participants (90%) reported having received some form of ACE-III training which consisted of either informal training (i.e. observing the practice of another clinician administering and scoring the ACE-III and then being observed administering and scoring the test themselves) or more formal training (i.e. through the NHS Education for Scotland (NES) online training programme). During their participation in the study, only 44% of participants utilised the ACE-III scoring and administration guide. Analyses were completed to explore participants' deviation from the TS between each of the following groups:

(i) Participants who had received ACE-III training (n = 37) and those who had not (n = 3).

(ii) Participants who utilised the ACE-III scoring and administration guide during their participation in the study (n = 18) and those who did not (n = 22).

(iii) Participants who reported using the ACE-III in clinical practice frequently (i.e. between once a week to once a month) (n = 36) and those who reported using the ACE-III infrequently (i.e. between once every three months to less than once every six months) (n = 4).

(iv) Participants who reported using the ACE-III in clinical practice for less than 5 years (n = 13) and those who reported using the ACE-III for 5 years or more (n = 27).

Independent-Samples Mann-Whitney U Tests were conducted to explore participants' deviation from the TS for groups (ii) and (iv). No significant differences in scoring performance were found. Table 5 presents the median deviation from the true score, interquartile range, and test statistics. Due to the very small number of participants for the sub-groups in (i) and (iii), only measures of central tendency are provided.

Table 5 Median deviation from the true score for participants based on training, whether the scoring guide was used, frequency of ACE III use, and length of time using the ACE III.

				Test Statistics	
	Mdn	*Interquartile Range*	*N*	*U*	*P*
Training					
Yes	1	1 - 3	37		
No	3	2 - 3	3		
Scoring guide used					
Yes	1	1 – 2.25	18	184.5	0.706
No	2	0 - 3	22		
Frequency of use					
Once a week - Once a month	1	1 - 3	36		
Once every 3 months – Less than once every 6 months	3.5	2.25 – 4.75	4		
Length of use					
Less than 5 years	1	0.5 – 1.5	13	125.0	0.134
5 years or more	2	2 - 2	27		

Calculation Errors

At session one, 10% (4/40) of participants made a calculation error while 'summing up' items to obtain an overall total score. This meant that for these ACE-III forms, the total score recorded differed from the actual summed total of items. Calculation errors did not exceed two points for any participant. At session two, only 3% (1/37) of participants made a calculation error affecting the overall total ACE-III score. The calculation error here consisted of one point.

Discussion

This study found that the *inter*- and *intra*-rater scoring accuracy of the ACE-III is generally good, with error mainly accounted for by the domains which require more subjective judgements, namely, the visuospatial and language tests. The majority of participants' scores on the ACE-III did not deviate from the TS by any more than 2 points at total, domain, and sub-domain levels. This margin of error should be acceptable in clinical practice given that the ACE-III is a screening measure rather than a diagnostic tool, and scores should be interpreted accordingly. However, if 'cut off' scores are adhered to rigidly, even small variations in scores could have significant clinical implications. These findings are in keeping with a previous study examining the scoring accuracy of the ACE-III's predecessor, the ACE-R, using similar methods (Crawford, 2010). Also in relation to ACE-III total scores, there was a tendency towards over-scoring across both sessions. It could be possible that some health professionals give patients 'the benefit of the doubt' when scoring screening tools, which may account for this finding. If patient performance on the ACE-III is over-scored in clinical practice this could potentially lead to false negative results and compromise the accurate and early identification of those individuals with a dementia.

The study also found that particular ACE-III domains were associated with less scoring accuracy, and highlighted the challenges of scoring where subjective judgements are required, as with the visuospatial and language tests. The study found a statistically significant difference between participant scores and the TS on the visuospatial and language domains, with large effect sizes. Deviation from the TS on these domains was mainly accounted for by scoring errors on the clock drawing item and the single word repetition item respectively, and this was observed across both sessions. In relation to the clock drawing item, participants may not have been aware of the guidelines for assigning points according to the clock face, the inclusion and distribution of numbers, and the placement of the clock hands, as outlined in the ACE-III scoring guide. The simulated older adult patient in the vignette did not perform very well on this test which may have increased the amount of subjectivity required in the process of scoring. The single word repetition item requires the patient to repeat several words correctly. In the vignette, the simulated older adult patient pronounced one of the words incorrectly at the first attempt, and then correctly upon a second attempt. It would seem that most of the participants gave this response full marks despite the scoring guide instructions stipulating that 'only the first attempt is scored'. This may indicate that some participants were not aware of the scoring instructions for this item.

Indeed, if individuals do not frequently refer to, and are not familiar with, the ACE-III administration and scoring guide, they may develop their own unique scoring methods. To improve scoring accuracy within these subjective domains, it may be helpful to include more detailed guidance on acceptable responses for these items within the ACE-III administration and scoring guidelines. That said, during their participation in the study, only 44% of participants utilised the ACE-III scoring and administration guide which may suggest that the guidelines are not routinely referred to in clinical practice. An alternative way of reducing scoring errors, particularly within the domains where a greater degree of rater interpretation is required, may be to adapt ACE-III scoring forms to include more detailed scoring information for these items on the form. Similar findings were highlighted by Crawford (2010) in relation to the ACE-R and, to address some of the issues, an online training programme was later developed by NES to help staff administer the newly developed ACE-III. The online training programme provides NHS staff with the knowledge and skills to use the ACE-III, giving them the opportunity to become familiar with administering and scoring the tool prior to clinical experience or as a refresher for those who have not administered and scored it for a significant length of time (NES, 2019). Indeed, perhaps the design of the current study could be adapted in future to examine the impact of viewing an instructional video on the ACE-III on scoring accuracy.

As noted previously, there was slightly less deviation from the ACE-III total TS at session two in comparison to session one which may be accounted for by practice effects (i.e. recall effects), particularly given that there was a relatively short time period of two months between session one and session two. The correlation coefficient obtained for participant ACE-III total scores between session one and two suggests that the ACE-III has a good level of test-retest consistency. It should be noted however that test-retest consistency was lower for the sub-domains.

The present study also examined whether scoring accuracy was affected by participants' experience of using the ACE-III and/or whether they had received training on how to administer and score it, and found that neither led to statistically significant differences in scoring performance. The majority of participants in the study had indeed received training on how to administer and score the ACE-III and most of the participants also had five or more years of experience using it therefore we may deduce from this that the ACE-III, even when used by appropriately skilled individuals, is still susceptible to some rater error.

These results should be interpreted with caution however given that the sub-groups of participants in these analyses had small numbers.

Only a small number of participants made a calculation error whilst summing up the scores which suggests that participants took time and care over this. Indeed, it was observed at the sessions that participants would double or even triple check that they had summed the total and domain scores accurately. One participant did not sum items to provide domain and total scores however this only occurred at session one and is therefore likely to have been an oversight.

Study Limitations

The design of this study enabled scoring accuracy to be investigated separately from administration consistency through the use of a scripted vignette and pre-determined ACE-III 'true' scores. It is possible that the 'true' scores may not have been an accurate reflection of actual scores and this is a potential limitation of the study design. The likelihood of this however is small given that the 'true' scores had been rated and agreed upon independently by two experienced Clinical Psychologists. Another potential limitation of the study design relates to the inability of participants to seek clarification and/or ask the patient to repeat themselves if they did not hear the responses as would be possible in real life settings. Some of the scoring errors made by participants, therefore, may have been due to their inability to hear the simulated older adult patient's responses, and/or mishearing responses, on the vignette. It should also be noted that while statistical significance is reported in the study when measuring difference between participants scores and predetermined 'true' scores, from a clinical perspective, this is of limited value as the question of how many points deviation from the 'true' score is clinically acceptable cannot be answered statistically. The judgement made that a few points deviation is likely to be clinically acceptable is based upon expert opinion rather than statistical assessment. In addition, all of the participants scores of the vignette correctly identified the patient as being in the impaired range. Indeed, the ACE-III 'true' score was significantly below the cut-off required for a diagnosis of dementia. If the 'true' score had been closer to the cut-off, that is, the point at which an individual's score moves from being within the 'normal' range to being within the 'impaired' range, it may have enhanced our understanding of scoring accuracy when an individual's presentation is less clear. This is, after all, the main aim of a cognitive screening test in clinical practice. As such, this is another limitation of the current research. A further limitation of the study relates to the test re-test method of assessing the scoring consistency of participants across

two different time points as there is potential for learning or recall effects. The short time interval of two months between the two test administrations may make these carryover effects more likely. To reduce these effects, a longer period of time between sessions would have been favourable. Finally, although the overall sample size in the study exceeded the power calculation requirements, analysis involving sub-groups of participants had small numbers and therefore may have influenced the validity of results.

Conclusion

Given the widespread use of cognitive screening tests in the assessment of dementia, there has been surprisingly little research on their scoring accuracy in practical settings. This study utilised a design which allowed for the scoring accuracy of the ACE-III to be explored independent of test administration. Findings suggest that the *inter*- and *intra*-rater scoring accuracy of the ACE-III is generally good, with error mainly accounted for by the domains which require more subjective judgements. Health professionals should therefore take these findings into account when scoring the ACE-III, and perhaps utilise the ACE-III administration and scoring guide to help improve accuracy. It may also be helpful to adapt ACE-III scoring forms to include more detailed scoring information for the items where a greater degree of rater interpretation is required. Health professionals may also wish to refresh their skills in administering and scoring the ACE-III through the NES online trainer programme. Indeed, if all health professionals score the ACE-III in a consistent manner, the accurate and early identification and management of those individuals with a dementia will be improved.

References

Alzheimer's Society (2018a). What is dementia? Available from: https://www.alzheimers.org.uk/about-dementia/types-dementia/what-dementia

Crawford, S. (2010). An Investigation of the Reliability of the Addenbrooke's Cognitive Examination – Revised (ACE-R). Available from: http://theses.gla.ac.uk/2142/1/2010CrawfordDClinPsy.pdf

Cullen, B., O'Neill, B., Evans, J.J., Coen, R.F. & Lawlor, B.A. (2007). A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry, 78*, 790-799.

Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral & biomedical sciences. *Behav Res Methods, 39*, 175-191.

Fleiss, J.L., Levin, B. & Paik, M.C. (2003) Statistical methods for rates and proportions, Hoboken, New Jersey: John Wiley & Sons, Inc.

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. *Dement Geriatr Cogn Disord, 36*, 242–50.

Kozora, E., Kongs, S., Hampton, M. & Zhang, L. (2008). Effects of examiner error on neuropsychological test results in a multi-site study. *Clin Neuropsychol, 22*, 977–88.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J. & Fischer, J. S. (2004). Neuropsychological assessment (4th ed.). New York, NY, US: Oxford University Press.

Machado, A., Baeta, E., Pimentel, P. & Peixoto, B. (2015). Psychometric and normative indictors of the Portuguese version of the Addenbrooke's cognitive examination-III. Preliminary study on a sample of healthy subjects. *Acta Neuropsychologica, 13*, (2), 127-136.

Milne, A., Culverwell, A., Guss, R., Tuppen, J. & Whelton, R. (2008). Screening for dementia in primary care: a review of the use, efficacy and quality of measures. *International Psychogeriatrics, 20*, 911-926.

National Institute for Health and Care Excellence (NICE). (2018). Dementia: assessment, management and support for people living with dementia and their carers. Available from: https://www.nice.org.uk/guidance/ng97/resources/dementia-assessment-management-and-support-for-people-living-with-dementia-and-their-carers-pdf-1837760199109

Newman, C.G.J., Bevins, A.D., Zajicek, J.P., Hodges, J.R., Vuillermoz, E., Dickenson, J.M., Kelly, D.S., Brown, S. & Noad, R.F. (2018). Improving the quality of cognitive screening assessments: ACEmobile, an iPad-based version of the Addenbrooke's Cognitive Examination-III. *Alzheimer's and Dementia, 10*, 182-187.

NHS Education for Scotland (NES). (2019). Education and Training: ACE-III Trainer. Available from:https://www.nes.scot.nhs.uk/education-and-training/by-discipline/psychology/multiprofessional-psychology/psychology-and-psychological-interventions-in-dementia/ace-iii-trainer.aspx

Noone, P. (2015). Questionnaire Review: Addenbrooke's Cognitive Examination-III. *Occupational Medicine, 65*, 418-420.

Scottish Government (2017). Scotland's National Dementia Strategy 2017-2020. Available from: http://www.gov.scot/Publications/2017/06/7735

Scottish Intercollegiate Guidelines Network (SIGN). (2006). Management of patients with dementia: A National Clinical Guideline. Available from: http://umh1946.edu.umh.es/wp-content/uploads/sites/172/2015/04/Management-of-patients-with-dementia-NHS.pdf

Stolwyk, R.J., O'Neill, M.H., McKay, A.J.D. & Wong, D.K. (2014). Are cognitive screening tools sensitive and specific enough for use after stroke? A systematic literature review. *Stroke, 45* (10), 3129-3134.

Sullivan, K. (2000) Examiners' errors on the Wechsler Memory Scale–Revised. *Psychol Rep 87*, 234–40.

Takenoshita, S., Terada, S., Yoshida, H., *et al.* (2019). Validation of Addenbrooke's cognitive examination III for detecting mild cognitive impairment and dementia in Japan. *BMC geriatrics*, *19* (1), 123.

Trochim, W.M.K. (2006). Web Center for Social Research Methods. Research Methods Knowledge Base (Reliability Section). Available from: http://www.socialresearchmethods.net/kb/reliable.php

Van Belle, G. & Arnold, A. (2000). Reliability of cognitive tests used in Alzheimer's disease. *Statistics in Medicine, 19*, 1411-1420.

Villarejo, A. & Puerta-Martin (2011). Usefulness of short tests in dementia screening. *Neurologia, 26* (7), 425-433.

Systematic Review Appendices

(Chapter 1)

Appendix 1.1 Publication Guidelines

AUTHOR GUIDELINES

1. SUBMISSION

Authors should kindly note that submission implies that the content has not been published or submitted for publication elsewhere except as a brief abstract in the proceedings of a scientific meeting or symposium.

Once the submission materials have been prepared in accordance with the Author Guidelines, manuscripts should be submitted online at https://mc.manuscriptcentral.com/gps.

Click here for more details on how to use ScholarOne.

Data Protection

By submitting a manuscript to, or reviewing for, this publication, your name, email address, institutional affiliation, and other contact details the publication might require, will be used for the regular operations of the publication, including, when necessary, sharing with the publisher (Wiley) and partners for production and publication. The publication and the publisher recognize the importance of protecting the personal information collected from users in the operation of these services, and have practices in place to ensure that steps are taken to maintain the security, integrity, and privacy of the personal data collected and processed. You can learn more at**https://authorservices.wiley.com/statements/data-protection-policy.html**.

Preprint Policy

The *International Journal of Geriatric Psychiatry* will consider for review articles previously available as preprints. Authors may also post the **submitted version** of a manuscript to a preprint server at any time. Authors are requested to update any pre-publication versions with a link to the final published article.

For help with submissions, please contact: **GPSeditorialoffice@wiley.com**

2. AIMS AND SCOPE

The rapidly increasing world population of aged people has led to a growing need to focus attention on the problems of mental disorder in late life. The aim of the International Journal of Geriatric Psychiatry is to communicate the results of original research in the causes, treatment and care of all forms of mental disorder which affect the elderly. The Journal is of interest to psychiatrists, psychologists, social scientists, nurses and others engaged in therapeutic professions, together with general neurobiological researchers.

The Journal provides an international perspective on the important issue of geriatric psychiatry, and contributions are published from countries throughout the world. Topics covered include epidemiology of mental disorders in old age, clinical aetiological research, post-mortem pathological and neurochemical studies, treatment trials and evaluation of geriatric psychiatry services.

Further information about the Journal, including links to the online sample copy and contents pages, can be found on the Journal homepage.

3. MANUSCRIPT CATEGORIES AND REQUIREMENTS

All tables, figures, supporting information and bibliographic entries must have a reference in the text. Tables and figures should be numbered (Figure 1, Table 1, Supplementary Figure 1). The journal prefers tables to be included in the main document after the reference list, each on an individual page alongside their legend. The journal prefers figures to be uploaded as individual files rather than included in the main document. However authors should indicate clearly where figures/tables should be inserted in the main document. All figures and tables must be accompanied by a legend. Word limits below include the main text and exclude tables, figure/table legends and references.

i. Research Articles

Research articles are the journal's primary mode of scientific communication. Peer-review of these will be handled by the most appropriate Editor.
Manuscript structure: Abstract (250 words max); Introduction; Materials and Methods; Results; Discussion; Acknowledgements; References; Tables; List of figure captions; List of supporting information legends. Overall combined limit of 6 figures/tables. Word limit: 3500 words.

ii. Review Article

The review articles published by the journal are characterised by the following: (i) high quality methodology; (ii) novelty in terms of the scope and content compared with other published reviews; (iii) coverage of an area of particular clinical and/or scientific importance; (iv) there being sufficient papers in the literature to make the review informative; and (v) the ability to make well supported conclusions of clinical and/or scientific importance. All reviews are subject to the same process of editorial and peer review as research articles.
Manuscript structure: Abstract (250 words max); Introduction; Content-appropriate headings; Acknowledgements; References; Tables; List of figure captions; List of supporting information legends. Overall combined limit of 6 figures/tables and 150 references. Word limit: 4500 words.

iii. Letter to the Editor

Letters to the Editor may be in response to issues arising from recently published articles, or short, free-standing pieces expressing an opinion.
Manuscript structure: No abstract; one continuous section with content-specific headings if needed. Overall combined limit of 1 figure/table. Word limit: 700 words.

iv. Commentary

Manuscript structure: No abstract; one continuous section with content-specific headings if needed. Word limit: 4500 words.

v. Editorial

Manuscript structure: No abstract; one continuous section with content-specific headings if needed. Word limit: 3500 words.

File formats

Text and tables should be in editable document formats, ideally .doc or .docx. Authors may also submit LaTeX files. On initial submission, LaTeX users should upload a single PDF file generated from source files and should state they are a LaTeX user in their cover letter. When submitting a revision, in addition to a single PDF file, please also submit all TeX source files as "Supplementary Material not for Review".

4. PREPARING YOUR SUBMISSION

Cover Letters

Cover letters are not mandatory; however, they may be supplied at the author's discretion.

Parts of the Manuscript

The manuscript should be submitted in separate files: main text file; figures.

Main Text File

The text file should be presented in the following order:

i. A short informative containing the major key words. The title should not contain abbreviations (see Wiley's **best practice SEO tips**);
ii. A short running title of less than 50 characters;
iii. The full names of the authors;
iv. The author's institutional affiliations where the work was conducted, with a footnote for the

author's present address if different from where the work was conducted;
v. The corresponding author details;
vi. The word count of the body text;
vii. Acknowledgments;
viii. The name(s) of any sponsor(s) of the research contained in the paper, along with grant number(s);
ix. Abstract, keywords and key-points;
x. Main text;
xi. References;
xii. Tables (each table complete with title and footnotes);
xiii. Figure legends;
xiv. Appendices (if relevant).

Figures and supporting information should be supplied as separate files.

Authorship

Please refer to the journal's Authorship policy in the Editorial Policies and Ethical Considerations section for details on author listing eligibility.

Acknowledgments

Contributions from anyone who does not meet the criteria for authorship should be listed, with permission from the contributor, in an Acknowledgments section. Financial and material support should also be mentioned. Thanks to anonymous reviewers are not appropriate.

Conflict of Interest Statement

Authors will be asked to provide a conflict of interest statement during the submission process. For details on what to include in this section, see the 'Conflict of Interest' section in the Editorial Policies and Ethical Considerations section below. Submitting authors should ensure they liaise with all co-authors to confirm agreement with the final statement.

Abstract

Abstracts are required for research and review articles. They must not exceed 250 words and should be divided into the following sections: 'Objectives', 'Methods'/'Design', 'Results', and 'Conclusions'.

Keywords

Please provide 3-10 keywords – these should be listed after the abstract in the main document and also entered into the submission system.

Key points

Please provide up to 4 key points – there should be listed after the keywords in the main document.

Main Text

Authors may submit using English UK or English US as spelling of accepted papers is converted during the production process, however this should be consistent throughout. See section 3 for information on manuscript types, word limits and other requirements. See Section 3: Manuscript categories and requirements for information on manuscript types, structure, word limit and other requirements.

References

All references should be numbered consecutively in order of appearance and should be as complete as possible. In text citations should cite references in consecutive order using Arabic superscript numerals.

For more information about this reference style, please see the AMA Manual of Style.

Reference examples follow:

Journal article

1. King VM, Armstrong DM, Apps R, Trott JR. Numerical aspects of pontine, lateral reticular, and inferior olivary projections to two paravermal cortical zones of the cat cerebellum. J Comp Neurol 1998;390:537-551.

Book

2. Voet D, Voet JG. Biochemistry. New York: John Wiley & Sons; 1990. 1223 p.

Please note that journal title abbreviations should conform to the practices of Chemical Abstracts.

Internet Document

1.

American Cancer Society. Cancer Facts & Figures 2003. http://www.cancer.org/downloads/STT/CAFF2003PWSecured.pdf. Accessed March 3, 2003.

Tables

Tables should be self-contained and complement, not duplicate, information contained in the text. They should be supplied as editable files, not pasted as images. Legends should be concise but comprehensive – the table, legend, and footnotes must be understandable without reference to the text. All abbreviations must be defined in footnotes. Footnote symbols: †, ‡, §, ¶, should be used (in that order) and *, **, *** should be reserved for P-values. Statistical measures such as SD or SEM should be identified in the headings.

Figure Legends

Legends should be concise but comprehensive – the figure and its legend must be understandable without reference to the text. Include definitions of any symbols used and define/explain all abbreviations and units of measurement.

Figures

Although authors are encouraged to send the highest-quality figures possible, for peer-review purposes, a wide variety of formats, sizes, and resolutions are accepted. **Click here** for the basic figure requirements for figures submitted with manuscripts for initial peer review, as well as the more detailed post-acceptance figure requirements.

Figures submitted as a colour original will appear in colour in the Journal's online edition free of charge. Colour figures will be printed in the Journal on the condition that authors contribute to the associated costs: £350 for the first page; £150 for each subsequent page thereafter. Corresponding authors will be invoiced post-publication.

Guidelines for Cover Submissions

If you would like to send suggestions for artwork related to your manuscript to be considered to appear on the cover of the journal, **please follow these general guidelines**.

Additional Files

Appendices

Appendices will be published after the references. For submission they should be supplied as separate files but referred to in the text.

Supporting Information

Supporting information is information that is not essential to the article, but provides greater depth and background. It is hosted online and appears without editing or typesetting. It may include tables, figures, videos, datasets, etc.

Click here for Wiley's FAQs on supporting information.

Note: if data, scripts, or other artefacts used to generate the analyses presented in the paper are available via a publicly available data repository, authors should include a reference to the location of the material within their paper.

General Style Points

The following points provide general advice on formatting and style.

• **Abbreviations:** In general, terms should not be abbreviated unless they are used repeatedly and the abbreviation is helpful to the reader. Initially, use the word in full, followed by the abbreviation in parentheses. Thereafter use the abbreviation only.
• **Units of measurement:** Measurements should be given in SI or SI-derived units. Visit the **Bureau International des Poids et Mesures (BIPM) website** for more information about SI units.
• **Numbers:** numbers under 10 are spelt out, except for: measurements with a unit (8mmol/l); age (6 weeks old), or lists with other numbers (11 dogs, 9 cats, 4 gerbils).
• **Trade Names:** Chemical substances should be referred to by the generic name only. Trade names should not be used. Drugs should be referred to by their generic names. If proprietary

drugs have been used in the study, refer to these by their generic name, mentioning the proprietary name and the name and location of the manufacturer in parentheses.

Article Preparation Support

Wiley Editing Services offers expert help with English Language Editing, as well as translation, manuscript formatting, figure illustration, figure formatting, and graphical abstract design – so you can submit your manuscript with confidence.

Also, check out our resources for **Preparing Your Article** for general guidance about writing and preparing your manuscript.

5. EDITORIAL POLICIES AND ETHICAL CONSIDERATIONS

Peer Review and Acceptance

The acceptance criteria for all papers is the quality and originality of the research and its significance to our readership. Except where otherwise stated, manuscripts are single-blind peer reviewed. Papers will only be sent to review if the Editor-in-Chief determines that the paper meets the appropriate quality and relevance requirements. Wiley's policy on confidentiality of the review process is **available here**.

Data Sharing and Data Accessibility

The *International Journal of Geriatric Psychiatry* recognizes the many benefits of archiving research data. The journal expects you to archive all the data from which your published results are derived in a public repository. The repository that you choose should offer you guaranteed preservation (see the registry of research data repositories at **https://www.re3data.org/**) and should help you make it findable, accessible, interoperable, and re-useable, according to FAIR Data Principles (**https://www.force11.org/group/fairgroup/fairprinciples**). All accepted manuscripts are required to publish a data availability statement to confirm the presence or absence of shared data. If you have shared data, this statement will describe how the data can be accessed, and include a persistent identifier (e.g., a DOI for the data, or an accession number) from the repository where you shared the data. Authors will be required to confirm adherence to the policy. If you cannot share the data described in your manuscript, for example for legal or ethical reasons, or do not intend to share the data then you must provide the appropriate data availability statement. The *International Journal of Geriatric Psychiatry* notes that FAIR data sharing allows for access to shared data under restrictions (e.g., to protect confidential or proprietary information) but notes that the FAIR principles encourage you to share data in ways that are as open as possible (but that can be as closed as necessary). Sample statements are available **here**. If published, all statements will be placed in the heading of your manuscript.

Data Citation

Please also cite the data you have shared, like you would cite other sources that your article refers to, in your references section. You should follow the format for your data citations laid out in the Joint Declaration of Data Citation Principles, **https://www.force11.org/datacitationprinciples**:

[dataset] Authors; Year; Dataset title; Data repository or archive; Version (if any); Persistent identifier (e.g. DOI)

For manuscripts reporting medical studies that involve human participants, a statement identifying the ethics committee that approved the study and confirmation that the study conforms to recognized standards is required, for example: **Declaration of Helsinki**; **US Federal Policy for the Protection of Human Subjects**; or **European Medicines Agency Guidelines for Good Clinical Practice**. It should also state clearly in the text that all persons gave their informed consent prior to their inclusion in the study.

Patient anonymity should be preserved. Photographs need to be cropped sufficiently to prevent human subjects being recognized (or an eye bar should be used). Images and information from individual participants will only be published where the authors have obtained the individual's free prior informed consent. Authors do not need to provide a copy of the consent form to the publisher; however, in signing the author license to publish, authors are required to confirm that consent has been obtained. Wiley has a **standard patient consent form available** for use.

Animal Studies

A statement indicating that the protocol and procedures employed were ethically reviewed and approved, and the name of the body giving approval, must be included in the Methods section of the manuscript. We encourage authors to adhere to animal research reporting standards, for example the **ARRIVE reporting guidelines** for reporting study design and statistical analysis; experimental procedures; experimental animals and housing and husbandry. Authors should also state whether experiments were performed in accordance with relevant institutional and national guidelines and regulations for the care and use of laboratory animals:

• US authors should cite compliance with the US National Research Council's **Guide for the Care and Use of Laboratory Animals**, the US Public Health Service's **Policy on Humane Care and Use of Laboratory Animals**, and **Guide for the Care and Use of Laboratory Animals**.
• UK authors should conform to UK legislation under the **Animals (Scientific Procedures) Act 1986 Amendment Regulations (SI 2012/3039)**.
• European authors outside the UK should conform to **Directive 2010/63/EU**.

Clinical Trial Registration

The journal requires that clinical trials are prospectively registered in a publicly accessible database and clinical trial registration numbers should be included in all papers that report their results. Authors are asked to include the name of the trial register and the clinical trial registration number at the end of the abstract. If the trial is not registered, or was registered retrospectively, the reasons for this should be explained.

Research Reporting Guidelines

Accurate and complete reporting enables readers to fully appraise research, replicate it, and use it. Authors are expected to adhere to recognised research reporting standards. The EQUATOR Network collects more than 370 reporting guidelines for many study types, including for:

• **Randomised trials**: **CONSORT**
• **Observational studies**: **STROBE**
• **Systematic reviews**: **PRISMA**
• **Case reports**: **CARE**
• **Qualitative research**: **SRQR**
• **Diagnostic / prognostic studies**: **STARD**
• **Quality improvement studies**: **SQUIRE**
• **Economic evaluations**: **CHEERS**
• **Animal pre-clinical studies**: **ARRIVE** • **Study protocols**: **SPIRIT**
• **Clinical practice guidelines**: **AGREE**

We also encourage authors to refer to and follow guidelines from:

• **Future of Research Communications and e-Scholarship (FORCE11)**
• **National Research Council's Institute for Laboratory Animal Research guidelines**
• **The Gold Standard Publication Checklist from Hooijmans and colleagues**
• **Minimum Information Guidelines from Diverse Bioscience Communities (MIBBI) website**
• **FAIRsharing website**

Species Names

Upon its first use in the title, abstract, and text, the common name of a species should be followed by the scientific name (genus, species, and authority) in parentheses. For well-known species, however, scientific names may be omitted from article titles. If no common name exists in English, only the scientific name should be used.

Genetic Nomenclature

Sequence variants should be described in the text and tables using both DNA and protein designations whenever appropriate. Sequence variant nomenclature must follow the current HGVS guidelines; see **http://varnomen.hgvs.org/**, where examples of acceptable nomenclature are provided.

Sequence Data

Nucleotide sequence data can be submitted in electronic form to any of the three major collaborative databases: DDBJ, EMBL, or GenBank. It is only necessary to submit to one database as data are exchanged between DDBJ, EMBL, and GenBank on a daily basis. The suggested wording for referring to accession-number information is: 'These sequence data have been submitted to the DDBJ/EMBL/GenBank databases under accession number U12345'. Addresses are as follows:

• DNA Data Bank of Japan (DDBJ): **http://www.ddbj.nig.ac.jp**

• EMBL Nucleotide Archive: **ebi.ac.uk/ena**
• GenBank: **www.ncbi.nlm.nih.gov/genbank**

Proteins sequence data should be submitted to either of the following repositories:

• Protein Information Resource (PIR): **pir.georgetown.edu**
• SWISS-PROT: **expasy.ch/sprot/sprot-top**

Conflict of Interest

The journal requires that all authors disclose any potential sources of conflict of interest. Any interest or relationship, financial or otherwise that might be perceived as influencing an author's objectivity is considered a potential source of conflict of interest. These must be disclosed when directly relevant or directly related to the work that the authors describe in their manuscript. Potential sources of conflict of interest include, but are not limited to: patent or stock ownership, membership of a company board of directors, membership of an advisory board or committee for a company, and consultancy for or receipt of speaker's fees from a company. The existence of a conflict of interest does not preclude publication. If the authors have no conflict of interest to declare, they must also state this at submission. It is the responsibility of the corresponding author to review this policy with all authors and collectively to disclose with the submission ALL pertinent commercial and other relationships.

Funding

Authors should list all funding sources in the Acknowledgments section. Authors are responsible for the accuracy of their funder designation. If in doubt, please check the Open Funder Registry for the correct nomenclature: **https://www.crossref.org/services/funder-registry/**

Authorship

The list of authors should accurately illustrate who contributed to the work and how. All those listed as authors should qualify for authorship according to the following criteria:

1. Have made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data;

2. Been involved in drafting the manuscript or revising it critically for important intellectual content;

3. Given final approval of the version to be published. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content; and

4. Agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Contributions from anyone who does not meet the criteria for authorship should be listed, with permission from the contributor, in an Acknowledgments section (for example, to recognize contributions from people who provided technical help, collation of data, writing assistance, acquisition of funding, or a department chairperson who provided general support). Prior to submitting the article all authors should agree on the order in which their names will be listed in the manuscript.

Additional Authorship Options. Joint first or senior authorship: In the case of joint first authorship, a footnote should be added to the author listing, e.g. 'X and Y should be considered joint first author' or 'X and Y should be considered joint senior author.'

Publication Ethics

This journal is a member of the **Committee on Publication Ethics (COPE)**. Note this journal uses iThenticate's CrossCheck software to detect instances of overlapping and similar text in

submitted manuscripts. Read Wiley's Top 10 Publishing Ethics Tips for Authors **here**. Wiley's Publication Ethics Guidelines can be found **here**.

ORCID

As part of the journal's commitment to supporting authors at every step of the publishing process, the journal encourages the submitting author (only) to provide an ORCID iD when submitting a manuscript. This takes around 2 minutes to complete. **Find more information here**.

6. AUTHOR LICENSING

If a paper is accepted for publication, the author identified as the formal corresponding author will receive an email prompting them to log in to Author Services, where via the Wiley Author Licensing Service (WALS) they will be required to complete a copyright license agreement on behalf of all authors of the paper.

Authors may choose to publish under the terms of the journal's standard copyright agreement, or **OnlineOpen** under the terms of a Creative Commons License.

General information regarding licensing and copyright is available **here**. To review the Creative Commons License options offered under OnlineOpen, please **click here**. (Note that certain funders mandate that a particular type of CC license has to be used; to check this please click **here.**)

Self-Archiving definitions and policies. Note that the journal's standard copyright agreement allows for self-archiving of different versions of the article under specific conditions. Please click here for more detailed information about self-archiving definitions and policies.

Open Access fees: If you choose to publish using OnlineOpen you will be charged a fee. A list of Article Publication Charges for Wiley journals is available **here**.

Funder Open Access: Please click **here** for more information on Wiley's compliance with specific Funder Open Access Policies.

7. PUBLICATION PROCESS AFTER ACCEPTANCE

Accepted article received in production

When your accepted article is received by Wiley's production production team, you (corresponding authors) will receive an email asking you to login or register with **Author Services.** You will be asked to sign a publication licence at this point.

Accepted Articles

The journal offers Wiley's Accepted Articles service for all manuscripts. This service ensures that accepted 'in press' manuscripts are published online shortly after acceptance, prior to copy-editing or typesetting. Accepted Articles are published online a few days after final acceptance and appear in PDF format only. They are given a Digital Object Identifier (DOI), which allows them to be cited and tracked and are indexed by PubMed. After the final version article is published (the article of record), the DOI remains valid and can still be used to cite and access the article.

Accepted Articles will be indexed by PubMed; submitting authors should therefore carefully check the names and affiliations of all authors provided in the cover page of the manuscript so it is accurate for indexing. Subsequently, the final copyedited and proofed articles will appear in an issue on Wiley Online Library; the link to the article in PubMed will update automatically.

Proofs

Once the paper is typeset, the author will receive an email notification with full instructions on how to provide proof corrections.

Please note that the author is responsible for all statements made in their work, including changes made during the editorial process – authors should check proofs carefully. Note that proofs should be returned within 48 hours from receipt of first proof.

Early View

The journal offers rapid publication via Wiley's Early View service. **Early View** (Online Version of Record) articles are published on Wiley Online Library before inclusion in an issue. Note there may be a delay after corrections are received before the article appears online, as Editors also need to review proofs. Once the article is published on Early View, no further

changes to the article are possible. The Early View article is fully citable and carries an online publication date and DOI for citations.

Citing this Article: eLocators

This journal now uses eLocators. eLocators are unique identifies for an article that service the same function page numbers have traditionally served in the print world. When citing this article, please insert the eLocator in place of the page number. For more information, please visit the Author Services eLocator page **here**.

8. POST PUBLICATION

Access and Sharing

When the article is published online:

• The author receives an email alert (if requested).
• The link to the published article can be shared through social media.
• The author will have free access to the paper (after accepting the Terms & Conditions of use, they can view the article).
• The corresponding author and co-authors can nominate up to ten colleagues to receive a publication alert and free online access to the article.

Print copies of an article can now be ordered (instructions are sent at proofing stage) at **www.sheridan.com/wiley/eoc**

Article Promotion Support

Wiley Editing Services offers professional video, design, and writing services to create shareable video abstracts, infographics, conference posters, lay summaries, and research news stories for your research – so you can help your research get the attention it deserves.

Measuring the Impact of an Article

Wiley also helps you measure the impact of your research through our specialist partnerships with **Kudos** and **Altmetric**.

9. EDITORIAL OFFICE CONTACT DETAILS

Editorial office: **GPSeditorialoffice@wiley.com**

Author Guidelines updated 25th November 2019

Appendix 1.2 Quality rating checklist

Section & Topic	No	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	**6**	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	**10a**	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined
RESULTS		
Participants	**19**	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

Appendix 1.3 Quality assessment scores for each included paper

	Almeida *et al.* (2019)	Brown *et al.* (2016)	Camargo *et al.* (2016)	Chen *et al.* (2013)	Dalrymple-Alford *et al.* (2010)	Federico *et al.* (2015)	Hendershott *et al.* (2019)	Hoops *et al.* (2009)	Kandiah *et al.* (2014)	Marras *et al.* (2013)	Ozdilek *et al.* (2014)	Sobreira *et al.* (2015)
Title or Abstract												
Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	2	2	2	2	2	2	2	2	2	2	2	1
Abstract												
Structured summary of study design, methods, results, and conclusions	2	2	2	2	2	2	2	2	2	2	2	2
Introduction												
Scientific and clinical background, including the intended use and clinical role of the index test	1	2	2	2	1	1	1	2	1	1	1	1
Study objectives and hypotheses	1	2	1	1	1	1	1	1	2	1	1	1
Methods												
Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	2	2	2	2	2	2	2	2	2	2	2	2
Eligibility criteria	2	2	1	2	2	2	1	1	1	2	1	2
On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	2	2	2	2	2	2	1	2	2	2	2	2
Where and when potentially eligible participants were identified (setting, location and dates)	2	1	1	2	2	0	2	2	2	2	2	2
Whether participants formed a consecutive, random or convenience series	2	2	2	2	2	2	2	2	2	2	2	2
Index test, in sufficient detail to allow replication	1	2	2	2	2	2	1	2	2	2	2	2
Reference standard, in sufficient detail to allow replication	1	2	2	2	2	2	1	2	2	2	2	2
Rationale for choosing the reference standard (if alternatives exist)	2	2	2	2	2	2	2	2	2	2	2	2
Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	0	0	0	0	0	0	1	0	1	0	1	0
Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	2	2	2	2	2	2	2	2	2	2	2	2
Whether clinical information and reference standard results were available to the performers/readers of the index test	0	0	0	0	0	2	0	0	0	2	0	0
Whether clinical information and index test results were available to the assessors of the reference standard	0	0	0	0	0	2	0	0	0	2	0	0

Methods for estimating or comparing measures of diagnostic accuracy	2	2	2	2	2	2	2	2	2	2	2	2
How indeterminate index test or reference standard results were handled	0	0	0	0	0	0	0	1	0	0	0	0
How missing data on the index test and reference standard were handled	0	0	0	0	0	0	0	0	0	1	0	0
Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	0	0	0	0	0	0	0	0	0	0	0	0
Intended sample size and how it was determined	1	0	0	0	0	0	0	0	0	0	0	0
Results												
Flow of participants, using a diagram	0	0	0	2	2	1	0	1	0	0	0	0
Baseline demographic and clinical characteristics of participants	2	2	2	2	2	2	2	2	2	2	2	2
Distribution of severity of disease in those with the target condition	1	1	1	1	1	1	1	1	1	1	1	1
Distribution of alternative diagnoses in those without the target condition	2	1	1	1	2	1	1	2	1	1	1	2
Time interval and any clinical interventions between index test and reference standard	0	0	0	2	2	0	0	2	0	2	2	2
Cross tabulation of the index test results (or their distribution) by the results of the reference standard	1	1	1	1	1	0	1	1	1	1	1	1
Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	1	1	2	2	2	1	2	2	1	2	2	1
Any adverse events from performing the index test or the reference standard	0	0	0	0	0	0	0	0	0	0	0	0
Discussion												
Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	1	0	1	2	2	1	0	2	1	0	1	0
Implications for practice, including the intended use and clinical role of the index test	1	2	2	2	2	2	2	2	2	2	2	2
Total quality score (max 62)	34	35	35	42	42	37	32	42	36	42	38	36

Major Research Project Appendices

(Chapter 2)

Appendix 2.1 Participant Information Sheet

Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University *of* Glasgow
Administration Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XH

An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

Information Sheet

I would like to invite you to take part in a research study. Before you decide whether or not you wish to take part, it is important for you to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. If there is anything that is not clear or if you would like more information, please contact me.

Who is conducting the research?

The research is being carried out by Hollie Thomson (Trainee Clinical Psychologist) from the Institute of Health and Wellbeing at the University *of* Glasgow.

What is the purpose of the study?

The study aims to explore how a clinical presentation is scored on the Addenbrooke's Cognitive Examination-III (ACE-III) by professionals working within the NHS.

Why have I been invited?

You have been invited to take part in this study because you have had experience administering and scoring the Addenbrooke's Cognitive Examination-III (ACE-III) clinically.

Do I have to take part?

No. Taking part in this research is entirely voluntary. It is up to you to decide whether or not you wish to take part. If you do decide to take part, you will be given a copy of this information sheet to keep and asked to sign a consent form to show that you have agreed to take part. You are free to withdraw at any time, without giving a reason.

What does taking part involve?

Taking part involves attending two sessions. The study will be conducted in meeting or lecture rooms at the University *of* Glasgow premises, or within NHS settings. At session one, you will be shown a vignette of an older adult actor being administered the ACE-III. You will be required to watch the vignette in its entirety whilst concurrently scoring an accompanying ACE-III form. You will also be asked to complete an additional information sheet detailing your occupation and your experience to date using the ACE-III as well as any formal training you have completed on the ACE-III. Session two will take place approximately two months after session one. At session two, you will again be asked to watch a vignette of an older adult actor being administered the ACE-III and again asked to score an accompanying ACE-III form. It is anticipated that each session should last no more than one hour.

What happens to the information?

Your identity and personal information will be completely confidential and seen only by researchers and regulators whose job it is to check the work of researchers. The information obtained will remain confidential and stored in a locked filing cabinet within the University *of* Glasgow premises and/or electronically on a password protected computer. The data is held in accordance with the Data Protection Act (2018), which means that it is kept safely and cannot be revealed to other people without your permission. The data will be stored in archiving facilities for up to 10 years in accordance with relevant Data Protection policies and regulations. After this period, further retention may be agreed or your data will be securely destroyed in accordance with the relevant standard procedures. The results of the study will be disseminated via email to the Community Mental Health Team's from which staff participated. Planned dissemination also includes completion of a DClinPsy thesis, publication in a scientific journal, and presentation at a scientific conference.

What are the possible benefits of taking part?

It is hoped that by taking part in this research, you will be providing valuable information regarding how professionals score the ACE-III in clinical situations. It is hoped that this information will influence further research into how dementia screening tools are utilised by clinicians.

Who has reviewed the study?

This study has been reviewed by the University *of* Glasgow College of Medical, Veterinary and life Sciences (MVLS) Ethics Committee and the NHS Greater Glasgow and Clyde (GG&C) Research and Development Department.

If you have any further questions?

If you would like further information about this research project please contact the researcher, Hollie Thomson, or alternatively her academic supervisor Prof. Jonathan Evans,

or her clinical supervisor, Dr. Stephanie Crawford. If you wish to seek general advice about participating in this study from someone **not** closely linked to the study, please contact Prof. Tom McMillan. Please find all contact details below.

Miss Hollie Thomson
Trainee Clinical Psychologist
Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University *of* Glasgow
Administration Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XH
Tel: 0141 211 0607
Email: Hollie.Thomson@ggc.scot.nhs.uk

Professor Jonathan Evans
Professor of Applied Neuropsychology
Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University *of* Glasgow
Administration Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XH
Tel: 0141 211 0694
Email: Jonathan.Evans@glasgow.ac.uk

Dr. Stephanie Crawford
Consultant Clinical Psychologist
Inverclyde Older People CMHT
Crown House
30 King Street
Greenock
PA15 1NL
Tel: 01475 558045
Email: Stephanie.Crawford@ggc.scot.nhs.uk

Professor Tom McMillan
Professor of Clinical Neuropsychology
Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University *of* Glasgow
Administration Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow

G12 0XH
Tel: 0141 211 0354
Email: Thomas.McMillan@glasgow.ac.uk

If you have a complaint about any aspect of the study?

If you are unhappy about any aspect of the study and wish to make a complaint, please contact the researcher in the first instance. The normal NHS complaint mechanisms are also available to you. The contact details are as follows:

Complaints Department
West Glasgow Ambulatory Care Hospital
Dalnair Street
Glasgow
G3 8SJ

Phone: 0141 201 4500

Email: complaints@ggc.scot.nhs.uk

Thank you for your time.

Appendix 2.2 Additional Information Sheet

Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University *of* Glasgow
Administration Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XH

An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

Additional Information Sheet

Please answer the following questions.

1. What is your occupation and how long have you been in this post?

---

2. How often do you use the Addenbrooke's Cognitive Examination-III (ACE-III) in your clinical work? (Please circle the most appropriate response).

   Once a week Once a fortnight Once a month

   Once every 3 months Once every 6 months Less than once every 6 months

3. (a) Have you participated in ACE-III training? (Please circle).

   Yes No

   (b) If you answered 'No' to question 3(a) please state briefly how you learned to use the ACE-III.

---

4. Please give an estimate of how long you have used the ACE-III in your clinical practice:

   _____years and _____months

Appendix 2.3 Participant Instructions

An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

Participant Instructions

Having read the information sheet, does anyone have any questions?

If you are happy to proceed and participate in this study could you please complete the consent form and the additional information sheets that are in your research packs. During the study you will be shown a vignette whereby the ACE-III is administered to a mock patient. The vignette will be shown in its entirety only once; you have to score each section of the ACE-III as you watch the vignette as though you are the one administering it. There should be no conferring throughout the study. There will be time to total up each of the sections at the end. A scoring guide is available on request should you wish to refer to it when scoring up at the end. In your pack, there is a sheet that provides the correct answers for the orientation section of the ACE-III for this vignette; please refer to this when scoring this section.

Appendix 2.4 Vignette Orientation Information

An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

Orientation Information

ACE-III: English Version A

Day: Thursday,

Date: 7th

Month: March

Year: 2013

Season: Spring

Floor: 4th

Hospital: Glasgow Royal Infirmary Hospital

Town: Glasgow

County: City of Glasgow

Country: Scotland

Current Prime Minister: David Cameron

Current President: Barack Obama

Appendix 2.5 Participant Consent Form

Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University *of* Glasgow
Administration Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XH

An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

Consent Form

Please initial box

I confirm that I have read and understood the Participant Information Sheet version 1, dated 01/02/2019.	☐
I confirm that I have read and understood the Privacy Notice version 1, dated 01/02/2019.	☐
I have had the opportunity to think about the information and ask questions, and understand the answers I have been given.	☐
I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected.	☐
I confirm that I agree to the way my data will be collected and processed and that data will be stored for up to ten years in accordance with relevant Data Protection policies and regulations.	☐
I understand that all data and information I provide will be kept confidential and will be seen only by study researchers and regulators whose job it is to check the work of researchers.	☐
I agree to take part in the study.	☐

Name of Participant (PRINT)	Date	Signature
Name of Researcher (PRINT)	Date	Signature

Appendix 2.6 MVLS College Ethics Committee Approval Letter

Dear Professor Jonathan Evans

MVLS College Ethics Committee
Project Title: *An Investigation into the Reliability of the Addenbrookes Cognitive Examination-III (ACE-III)*

Project No: 200180141

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. We are happy therefore to approve the project, subject to the following conditions.

- Project end date as stipulated in original application.
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (http://www.gla.ac.uk/media/media_227599_en.pdf)
- The research should be carried out only on the sites, and/or groups defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- For projects requiring the use of an online questionnaire, the University has an Online Surveys account for research. To request access, see the University's application procedure at https://www.gla.ac.uk/research/strategy/ourpolicies/useofonlinesurveystoolforresearch/.
- You should submit a short end of study report to the Ethics Committee within 3 months of completion.
- Permissions from clinical leads will also be required for this study of NHS staff.

Yours sincerely

[redacted]

Dr Terry Quinn

Appendix 2.7 NHS GG&C R&D Approval Letter

Administrator: Mrs Elaine O'Neill
Telephone Number: 0141 232 1815
E-Mail: elaine.o'neill2@ggc.scot.nhs.uk
Website: www.nhsggc.org.uk/r&d

R&D Management Office
West Glasgow ACH
Dalnair Street
Glasgow G3 8SW

26 April 2019

Miss Hollie Thomson
Ins of Health and Wellbeing
Admin Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow G12 0XH

NHS GG&C Board Approval

Dear Miss H Thomson,

Study Title:	**An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)**
Principal Investigator:	**Miss Hollie Thomson**
GG&C HB site	**NHS GG&C Older People's Community Mental Health**
Sponsor	**NHS Greater Glasgow and Clyde**
R&D reference:	**GN19MH124**
REC reference:	**n/a**
Protocol no:	**V1; 01/02/19**

I am pleased to confirm that Greater Glasgow & Clyde Health Board is now able to grant **Approval** for the above study.

Conditions of Approval

1. **For Clinical Trials** as defined by the Medicines for Human Use Clinical Trial Regulations, 2004

    a. During the life span of the study GGHB requires the following information relating to this site
        i. Notification of any potential serious breaches.
        ii. Notification of any regulatory inspections.

It is your responsibility to ensure that all staff involved in the study at this site have the appropriate GCP training according to the GGHB GCP policy (www.nhsggc.org.uk/content/default.asp?page=s1411), evidence of such training to be filed in the site file.

2. **For all studies** the following information is required during their lifespan.
   a. Recruitment Numbers on a monthly basis
   b. Any change of staff named on the original SSI form
   c. Any amendments – Substantial or Non Substantial
   d. Notification of Trial/study end including final recruitment figures
   e. Final Report & Copies of Publications/Abstracts

Please add this approval to your study file as this letter may be subject to audit and monitoring.

Your personal information will be held on a secure national web-based NHS database.

I wish you every success with this research study

Yours sincerely,

Mrs Elaine O'Neill
Senior Research Administrator

Cc: Miss Emma Jane Gault (Glasgow University)
Prof Jonathan Evans (Glasgow University)

Appendix 2.8 Major Research Project Proposal

Doctorate in Clinical Psychology

MRP Proposal: Version

Date of Submission: 28.09.18

Maximum Word Count: 3,000

Actual Word Count: 3,321

An Investigation into the Reliability of the Addenbrooke's Cognitive Examination-III (ACE-III)

Abstract

Background: Dementia is a leading cause of disability amongst older adults in the developed world. In recent years, an increasing emphasis has been placed on the early detection and diagnosis of dementia, and timely access to post-diagnostic support. Cognitive screening tests are essential tools in facilitating the process of early detection and dementia diagnosis and are currently widely used in clinical practice. The Addenbrooke's Cognitive Examination-III (ACE-III) is one such tool; a cognitive screening test commonly used in the assessment of dementia. Despite it being recommended in several evidence-based guidelines and being widely used in the NHS, the rater reliability of the ACE-III has never been formally evaluated. ***Aims:*** To investigate rater accuracy in scoring the ACE-III both in terms of its total and sub-category scores across different raters and by the same raters at two different time points. A secondary exploratory analysis examines whether scoring accuracy is affected by participants' experience of using the ACE-III and/or whether they have had formal training on how to administer and score the ACE-III. ***Methods:*** A filmed vignette of the ACE-III being administered to an older adult actor (mock patient) will be used to assess scoring accuracy across different raters. The vignette has a pre-determined 'true' score. Participants will be asked to view the filmed vignette whilst simultaneously completing an ACE-III scoring sheet. Following a two-month period, participants will be invited back to view and score the same vignette again to assess *intra*-rater reliability. Participants will be NHS staff working in Older People's Community Mental Health Teams and Memory Assessment Centres who routinely administer and score the ACE-III as part of their clinical practice. ***Applications:*** This study addresses a gap in the literature by providing information on the rater reliability of the ACE-III. Given that the ACE-III is a screening test commonly used in the assessment of dementia and has the potential to facilitate the early identification and management of those individuals with a dementia, establishing its reliability is a highly clinically relevant area for research.

Introduction

Background

Dementia is an umbrella term for a range of progressive conditions that are characterised by global impairment of cognitive functioning (Alzheimer's Society, 2018). Dementia is a leading cause of disability amongst older adults in the developed world, and constitutes one of the most significant, and costly, challenges currently facing health and social care services (Milne *et al*., 2008). In the UK alone, there are approximately 850,000 people currently living with a dementia (Alzheimer's Society, 2018). Given that age is the biggest risk factor for developing a dementia, and that the average life expectancy is increasing, it is not surprising that the incidence and prevalence of dementia is expected to rise rapidly over the next several decades. Indeed, it is estimated that by 2051 there will be around two million people in the UK living with a dementia (Alzheimer's Society, 2018). Over the past decade, the Scottish Government has made dementia a public health priority and has set out a range of commitments and objectives to improve the quality of care for individuals with a dementia, with increasing emphasis being placed on early detection and diagnosis, and timely access to post-diagnostic support (Scottish Government, 2017). Early detection of dementia is essential to ensuring early intervention and management of the condition, and improving the individuals' quality of life. (Scottish Government, 2017).

The Role of cognitive screening tests in dementia diagnosis

The National Institute for Health and Care Excellence (NICE, 2018) has recently published up to date guidelines on the diagnosis and management of dementia which stipulate the use of a validated brief structured cognitive screening test as part of the initial stage of a comprehensive dementia assessment. Indeed, the assessment of cognitive functioning is central to the process of early detection and diagnosis of dementia and is therefore arguably one of the most important assessments made by clinicians in clinical practice (Machado *et al.,* 2015). In the context of time constraints, the efficient early detection of dementia requires the use of an objective cognitive screening test which is brief, reliable, easy to administer, and acceptable to patients (Cullen *et al*., 2007; Villarejo & Puerta-Martin, 2011). It is not surprising then that in recent years there has been an increase in the number of dementia screening tests proposed and studied.

The main aim of a cognitive screening test is to provide information about the presence, or absence, of a cognitive impairment (Lezak *et al*., 2004). This is often inferred from the individuals' score on the test compared to referenced norms (Cullen *et al.,* 2007). A central feature of most screening tests is the 'cut off' score, that is, the point at which an individual's score moves from being within the 'normal' range to being within the 'impaired' range, i.e. indicating the presence of a dementia. The success of a screening test in achieving this aim undoubtedly lies in its psychometric robustness (Cullen *et al.,* 2007). Cognitive screening tests should have good sensitivity and specificity to ensure that individuals with a cognitive impairment are not missed and that individuals without a cognitive impairment are not misidentified as impaired and referred on for further testing unnecessarily (Stolwyk *et al*., 2014).

One of the most important attributes of any assessment instrument is its reliability, that is, the extent to which the test produces consistent measurements when repeated under replicate conditions (Van Belle & Arnold, 2000). There are a number of factors which may influence the reliability of a test including individual factors (i.e. mood, fatigue, and motivation), environmental factors (i.e. temperature and noise), and rater factors (i.e. training, experience, judgement). These factors are sources of measurement error in the assessment process. If these measurement errors did not exist, one would expect that an individual would obtain the same test score, their 'true' score, each time they took the test. It is generally accepted however that the observed score on a test equates to the 'true' score plus some degree of measurement error (Trochim, 2006). Thus, the extent to which a cognitive test can minimise the impact of measurement errors, and ensure that the score obtained is as close to the individuals 'true' score as possible, is an indication of its reliability (Trochim, 2006). Inter-rater reliability is the level of scoring consistency between raters evaluating the same test. High levels of agreement between raters is thought to be indicative of high inter-rater reliability. However, agreement between raters does not necessarily equate to scoring accuracy when evaluating a test, since raters may show high levels of agreement but show low levels of accuracy. Therefore, rater accuracy in scoring tests across different raters is an important aspect of reliability that must be taken into consideration in inter-rater reliability studies. Intra-rater reliability assesses scoring consistency by the same raters at two different time points. High levels of scoring consistency by raters at two different time points, indicates high levels of intra-rater reliability.

The ACE-III as a cognitive screening test for dementia

The Addenbrooke's Cognitive Examination, which is now in its third edition (ACE-III), is one of the cognitive screening tests recommended by the Scottish Intercollegiate Guideline Network (SIGN) in its guideline on the Management of Patients with Dementia (SIGN, 2006). The ACE-III is comprised of five subscales, each representing a cognitive domain; Attention, Memory, Fluency, Language, and Visuo-spatial. A maximum score of 100 can be obtained, with higher scores indicative of better memory and cognitive performance. The ACE-III is designed to be sensitive to the early stages of dementia, and has been shown to compare favourably with its predecessor, the ACE- R, with similar levels of sensitivity and specificity in the assessment of cognitive deficits in people with Alzheimer's disease and fronto-temporal dementia (Hsieh *et al*., 2013). The ACE-III continues to show high sensitivity and specificity for dementia at cut offs previously recommended, i.e. 88 (sensitivity = 1.0; specificity = 0.96) and 82 (sensitivity = 0.93; specificity = 1.0). It takes approximately 15 minutes to administer the ACE-III and it therefore meets the requirements of a screening test which is time efficient (Cullen *et al.,* 2007). There are three versions of the ACE-III available to reduce any practice effects with repeat testing. Appropriate use of the ACE-III requires clinicians who are able to accurately administer, score, and interpret the test.

The current study

Although the psychometric properties of cognitive screening tests are widely reported, there is a dearth of research exploring the rater reliability and scoring accuracy of such tests among clinicians who routinely use them in clinical practice (Newman *et al.,* 2018). Previous studies examining test scoring on cognitive screening measures have found high rates of error (Crawford, 2010; Kozora, 2018; Sullivan, 2000) which highlights the importance of usability research, and ensuring that tests are used appropriately in real-world settings. Similarly, the consistency with which a test is scored by a range of clinicians is rarely reported. This is particularly relevant when there are items on a test which require more subjective scoring judgments.

Despite being recommended in several evidence-based guidelines as a screening test for dementia, and being widely used in the NHS, the rater reliability of the ACE-III has never been formally evaluated. As with any method of objective testing, it is essential to establish whether the ACE-III is accurately and consistently scored across different raters and by the

same raters at different time points. Indeed, this is highly relevant to ensuring the accurate and early identification and management of those individuals with a dementia. Since the clinical utility of a test is limited without consideration of reliability, this is a highly clinically relevant area for research.

Aims

This study aims to investigate rater accuracy in scoring the ACE-III. It will explore how accurately the ACE-III is scored both in terms of its total and sub-category scores across different raters and by the same raters at two different time points. It will also consider whether scoring accuracy is affected by participants' experience of using the ACE-III and/or whether they have had formal training on how to administer and score the ACE-III.

Research Questions

The specific research questions in relation to this study are:

(i) Do participants' ACE III total scores differ significantly from the ACE III 'true' score?

(ii) Do participants' ACE-III sub-category scores differ significantly from the ACE-III 'true' scores?

(iii) What is the test-retest scoring consistency across two different time points?

Secondary Exploratory Analysis

A secondary exploratory analysis will be conducted to examine whether scoring accuracy is affected by participants' experience of using the ACE-III and/or whether they have had formal training on how to administer and score the ACE-III.

Method

Participants

The participants in this study will be NHS staff working in Older People's Community Mental Health Teams and Memory Assessment Centres in Greater Glasgow and Clyde and Dudley and Walsall Mental Health Partnership Trust who routinely administer and score the

ACE-III as part of their clinical practice. To be considered eligible for inclusion in this study, participants are required to have administered and scored the ACE-III independently in clinical practice on at least one occasion with an older adult patient. Some participants may have received formal training on how to administer and score the ACE-III.

Justification of Sample Size

The power calculation was based on the primary method of data analysis; a one sample t-test. In a similar study examining the rater reliability of the Addenbrooke's Cognitive Examination-Revised (ACE-R), Crawford (2010) reported medium effect sizes. A medium effect size is deemed to be sufficient enough to detect a clinically significant difference, that is, a difference large enough to alter interpretation of the ACE-III results. G*Power (Faul *et al.*, 2007) indicated that a minimum of 34 participants would be required to detect a medium effect size, with p=0.05, and a power of 0.8.

Measures

The Addenbrooke's Cognitive Examination-III (ACE-III) is the primary outcome measure.

Design Procedure

Permission was given to use a filmed vignette from the online NHS Education for Scotland (NES) ACE-III Trainer programme for the purposes of this study. The University *of* Glasgow media services originally filmed and produced the vignette. The vignette is of a Clinical Psychologist administering the ACE-III to an older adult actor (mock patient). The vignette has pre-determined 'true' scores which had been rated, and agreed upon, by two experienced Clinical Psychologists. Pre-determined 'true' scores allows for scoring consistency to be investigated separately from administration consistency. A script for the vignette was developed by an advisory group of Clinical Psychologists with extensive neuropsychological experience within older adult services. The vignette has a 'true' total score of 63/100.

Research Procedure

Participants will be recruited from Older People's Community Mental Health Teams across NHS Greater Glasgow and Clyde. Potential participants will be invited to take part in the study via email. Participants who are interested in taking part will receive an information sheet about the study (Appendix 1.1) and a consent form (Appendix 1.2). Written consent will be obtained from all participants prior to the commencement of the study. Participants who opt in to the study will then be invited to a group session lasting approximately one hour, where they will watch a vignette whilst simultaneously completing an ACE-III scoring sheet. The vignette will be shown to participants in groups and played on a projector screen. Participants will be asked to view the vignette once and will not be permitted to pause or rewind it, reflecting actual clinical practice. The ACE-III administration and scoring guide will be made available to participants on request. Participants will not be permitted to consult each other regarding scoring and will be blinded from each other's ratings. Participants will also be asked not to discuss their scoring after the session. In addition to scoring the vignettes, participants will also be asked to complete an information form detailing their profession, ACE-III experience, and whether or not they have completed any formal training in using the ACE-III (Appendix 1.3). A standardised set of instructions will be given verbally to each group of participants prior to commencing the study (Appendix 1.4). Participants will also be given the correct orientation information for the vignette (Appendix 1.5). To evaluate *intra*-rater reliability, participants will be invited back after a period of approximately two months to view and score the same vignette again, under the same conditions and in an identical manner to the first time point. Participants will be blinded from their previous ratings. Participants will also be asked for an update as to whether they have received any additional ACE-III training since the previous test session.

Settings and Equipment

The study will be conducted in meeting or lecture rooms at the University *of* Glasgow premises, or within NHS settings. A laptop and projector will be required to allow participants to view the vignette. ACE-III scoring sheets, administration and scoring guides, and additional information sheets will also be required for each participant.

Data Analysis

Descriptive statistics will be presented. Proportions and standard deviations will be used to analyse the variation of test scores from the 'true' scores. Percentages will also be used to determine how many participants correctly identified the 'true' total and sub-test scores. Analysis will involve one sample t-tests to compare whether the participants' ACE-III total scores differed significantly from the predetermined ACE-III 'true' scores, and to investigate whether the results obtained differ significantly from the 'true' sub-test scores. Correlational methods will be used to explore *intra*-rater reliability, that is, whether participants' ACE-III scores are consistent across two different time points. Correlational methods will also be used to explore whether scoring accuracy is affected by participants' experience of using the ACE-III and/or whether they have had formal training on how to administer and score the ACE-III. Statistical analyses will be carried out using SPSS.

Ethical Considerations

Prior to commencement of the study, ethical approval will be sought from the College of Medical, Veterinary, and Life Sciences (MVLS) Ethics Committee. Approval will also be sought from the NHS Greater Glasgow and Clyde Research and Development (R&D) Management Service, and the Dudley and Walsall Health Research Authority (HRA). Practice will be informed by the British psychological Society Code of Ethics and Conduct (2018). Participants will be provided with information about the study prior to their participation to facilitate the process of informed consent. Participants will be required to give informed consent prior to their participation. Following participation in the study, participants will be informed about how the data will be used and reminded that individual results will not be revealed. All of the data gathered for the purposes of the study will remain confidential and will be stored securely in accordance with the General Data Protection Regulations (GDPR). It is not anticipated that the study will cause distress to participants.

Health and Safety Issues

Researcher and Participant Safety Issues

No significant health and safety issues are anticipated. Participants will be NHS staff and the study will be conducted at the University *of* Glasgow premises, or within NHS settings.

Financial Issues

No significant financial costs are expected.

Practical Applications

This study addresses a gap in the literature by providing information on the rater reliability of the ACE-III.

Timetable

Draft Proposal	29th June 2018
Final Proposal	28th September 2018
Obtain Ethical Approval	by March 2019
Data Collection	April to September 2019
Data Analysis and Write up	October to December 2020
Submission	28th February 2020

References

Alzheimer's Society (2018). *What is dementia?* Available from: https://www.alzheimers.org.uk/about-dementia/types-dementia/what-dementia

Crawford, S. (2010). *An Investigation of the Reliability of the Addenbrooke's Cognitive Examination – Revised (ACE-R).* Available from: http://theses.gla.ac.uk/2142/1/2010CrawfordDClinPsy.pdf

Cullen, B., O'Neill, B., Evans, J.J., Coen, R.F. & Lawlor, B.A. (2007). A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry, 78*, 790-799.

Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral & biomedical sciences. *Behav Res Methods, 39*, 175-191.

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. *Dement Geriatr Cogn Disord*, *36*, 242–50.

Kozora, E., Kongs, S., Hampton, M. & Zhang, L. (2008). Effects of examiner error on neuropsychological test results in a multi-site study. *Clin Neuropsychol, 22*, 977–88.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J. & Fischer, J. S. (2004). *Neuropsychological assessment* (4th ed.). New York, NY, US: Oxford University Press.

Machado, A., Baeta, E., Pimentel, P. & Peixoto, B. (2015). Psychometric and normative indictors of the Portuguese version of the Addenbrooke's cognitive examination-III. Preliminary study on a sample of healthy subjects. *Acta Neuropsychologica, 13*, (2), 127-136.

Milne, A., Culverwell, A., Guss, R., Tuppen, J. & Whelton, R. (2008). Screening for dementia in primary care: a review of the use, efficacy and quality of measures. *International Psychogeriatrics, 20,* 911-926.

National Institute for Health and Care Excellence (NICE). (2018). *Dementia: assessment, management and support for people living with dementia and their carers.* Available from: https://www.nice.org.uk/guidance/ng97/resources/dementia-assessment-management-and-support-for-people-living-with-dementia-and-their-carers-pdf-1837760199109

Newman, C.G.J., Bevins, A.D., Zajicek, J.P., Hodges, J.R., Vuillermoz, E., Dickenson, J.M., Kelly, D.S., Brown, S. & Noad, R.F. (2018). Improving the quality of cognitive screening assessments: ACEmobile, an iPad-based version of the Addenbrooke's Cognitive Examination-III. *Alzheimer's and Dementia, 10,* 182-187.

Noone, P. (2015). Questionnaire Review: Addenbrooke's Cognitive Examination-III. *Occupational Medicine, 65*, 418-420.

Scottish Government (2017). *Scotland's National Dementia Strategy 2017-2020*. Available from: http://www.gov.scot/Publications/2017/06/7735

Scottish Intercollegiate Guidelines Network (SIGN). (2006). *Management of patients with dementia: A National Clinical Guideline.* Available from: http://umh1946.edu.umh.es/wp-content/uploads/sites/172/2015/04/Management-of-patients-with-dementia-NHS.pdf

Stolwyk, R.J., O'Neill, M.H., McKay, A.J.D. & Wong, D.K. (2014). Are cognitive screening tools sensitive and specific enough for use after stroke? A systematic literature review. *Stroke, 45* (10), 3129-3134.

Sullivan, K. (2000) Examiners' errors on the Wechsler Memory Scale–Revised. *Psychol Rep 87*, 234–40.

Trochim, W.M.K. (2006). Web Center for Social Research Methods. Research Methods Knowledge Base (Reliability Section). Available from: http://www.socialresearchmethods.net/kb/reliable.php

Van Belle, G. & Arnold, A. (2000). Reliability of cognitive tests used in Alzheimer's disease. *Statistics in Medicine, 19*, 1411-1420.

Villarejo, A. & Puerta-Martin (2011). Usefulness of short tests in dementia screening. *Neurologia, 26* (7), 425-433.