An Evaluation of Brief Cognitive Tests for the Identification of Mild Cognitive Impairment

Seline Nezihe Ozer

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Psychology

August 2016

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 2 of the thesis has appeared in publication as follows:

Ozer S, Young J, Champ C & Burke M. A systematic review of the diagnostic test accuracy of brief cognitive tests to detect amnestic mild cognitive impairment. International Journal of Geriatric Psychiatry. 2016. Advance online publication, doi: 10.1002/gps.4444.

I was responsible for designing the systematic review, selecting studies, extracting data from included studies, running the analyses and writing the article. C Champ was a second reviewer and assisted with selecting studies and extracting data from included studies. J Young and M Burke provided advice and guidance and commented on drafts.

The work in Chapter 4 of the thesis has appeared in publication as follows:

Ozer S, Noonan K, Burke M, Young J, Barber S, Forster A, Jones R. The validity of the Memory Alteration Test and the Test Your Memory test for community-based identification of amnestic mild cognitive impairment. Alzheimer's & Dementia. 2016. Advance online publication, doi:10.1016/j.jalz.2016.03.014

I was responsible for managing the study and was involved in collecting data. I was also responsible for running the statistical analyses and writing the article. All authors (except M Burke) were co-applicants on the NIHR grant application that funded the study. K Noonan was responsible for designing the cognitive assessment testing schedule, provided advice and commented on drafts. M Burke provided advice and

commented on drafts. J Young was chief investigator on the study, provided advice and commented on drafts. S Barber conceived the original idea for the study, provided advice and commented on drafts. A Forster was a senior advisor on the study and commented on drafts. R Jones provided advice and guidance on the design of the study.

Acknowledgements

First and foremost I would like to give thanks to the participants who generously donated their time to take part in the study. I would also like give huge thanks my supervisors Dr Melanie Burke and Prof John Young, who have provided much help, support and guidance throughout the PhD process and have given invaluable feedback on drafts. The M@T and TYM validity study was a team effort and I must give thanks to and acknowledge the expertise and input from the whole grant co-applicant team including Dr Krist Noonan, Prof Roy Jones, Dr Sally Barber, Prof Anne Forster, Dr Jane Kirby and Prof John Young, as well as the team of researchers who were involved with recruitment and data collection (Gill, Sarah, Lesley, Rosie, Alex, Katie, Nicola and Jaime). There was a huge amount of data collected for the study and I must also give thanks to Chung Fu for designing a very user friendly data entry system and to Dr Tom Crocker for managing the randomisation process. In addition I would like to thank Claire Champ for her role as second reviewer on the systematic review, as well as Dr Sarah Bauermeister and Prof David Bunce for their advice on the design and analysis of the reaction time study.

Finally I would like to thank my partner Simon, my parents and family for their love and support throughout the process and always, as well as my colleagues at the Academic Unit of Elderly Care and Rehabilitation for their moral support, particularly fellow PhD students Jess, Faye, Nicola and Jen for sharing the pain of completing a PhD!

Abstract

Mild cognitive impairment (MCI), a term used to describe the transitional state between normal aging and established dementia, has been identified as a potentially effective time point at which to target interventions to prevent or slow the decline into dementia. An efficient way of identifying people with MCI is required as a first step in conducting large scale studies to develop and evaluate interventions targeted at this high risk population.

The aim of the thesis was to investigate the effectiveness of a range of brief cognitive tests to be used for the purpose of identifying people with MCI in an efficient and accurate manner. A systematic review of the literature found that over 40 brief cognitive tests have been developed and tested to identify amnestic MCI (aMCI). However, the majority of these previous studies were conducted in secondary care settings and were at high risk of unblinding the assessment process, which may have exaggerated the diagnostic accuracy of the assessed tests.

The Memory Alteration Test (M@T) and Test Your Memory test (TYM) were selected as potentially useful brief cognitive tests for aMCI and their validity was assessed within a cohort of older people recruited from the community, without prior knowledge of their cognitive status. A total of 472 older people were assessed for MCI according to the Petersen criteria using a standardised battery of neuropsychological tests. A prevalence of MCI of 16.5% was found within the assessed cohort and the M@T was found to be more accurate than the TYM at detecting aMCI, performing with a sensitivity of 85% and specificity of 84%.

The ability of reaction time task derived measures to identify cognitive impairment within the cohort was also assessed. These were not as accurate at predicting cognitive status as the M@T. However, they did demonstrate some promising discriminative abilities and should be further explored as potentially useful alternatives to the traditionally used memory tests, which may be influenced by administrator bias, education level and language ability.

Contents

List of Tables

List of Figures

Abbreviations

15-OT:	15-Objects Test
AACD:	Aging-Associated Cognitive Decline
AAMI:	Age-Associated Memory Impairment
AD:	Alzheimer's disease
aMCI:	Amnestic Mild Cognitive Impairment
AQT-CF:	A Quick Test of Cognitive Speed (colour-form naming)
AUC:	Area Under ROC Curve
AVLT-SR:	Auditory Verbal Learning Test- Short Delay Recall
BNT-M:	Boston Naming Test-Modified
CDT:	Clock Drawing Test
CDR:	Clinical Dementia Rating scale
CERAD:	Consortium to Establish a Registry for Alzheimer's Disease
CERAD-WLDR:	CERAD- Word List Memory Test Delayed Recall
CERAD-WLLE:	CERAD- Word List Memory Test Wordlist Learning
CERAD-WLREDI:	CERAD-WL Recognition Discrimination Index
CERAD-WLRE:	CERAD- Wordlist Memory Test Recognition
CERAD-WLSA:	CERAD- Wordlist Memory Test Savings
CI:	Confidence Interval
CIND:	Cognitively Impaired, Not Demented
DTA:	Diagnostic Test Accuracy
ECR:	Enhanced Cued Recall
FBMS:	Florida Brief Memory Screen
FCSRT:	Free and Cued Selective Recall Reminding Test
FN:	False Negatives
FP:	False Positives

GDS:	Geriatric Depression Scale / Global Deterioration Scale
GNT:	Graded Naming Test
HDS-R:	Hasegawa's Dementia Scale-Revised
HVLT-LE:	Hopkins Verbal Learning Test- Wordlist Learning
ICC:	Intraclass Correlation Coefficient
LR+:	Postive Likelihood Ratio
LR-:	Negative Likelihood Ratio
M@T:	Memory Alteration Test
MCI:	Mild Cognitive Impairment
MES:	Memory and Executive Screening
MIS-plus:	Memory Impairment Screen Plus (Dutch version of MIS)
MMSE:	Mini Mental State Examination
MoCA:	Montreal Cognitive Assessment
naMCI:	Non-amnestic MCI
NPV:	Negative Predictive Value
PPV:	Positive Predictive Value
ROC:	Receiver Operating Characteristic curve
TMT:	Trail Making Test
TN:	True Negatives
TP:	True Positives
TYM:	Test Your Memory test
VAT:	Visual Association Test
VFT:	Verbal Fluency Task
VLT:	Verbal Learning Task
WAIS:	Wechsler Adult Intelligence Scale

Chapter 1: Introduction

As the population ages, dementia is becoming a common and increasing problem for individuals, families and society. Dementia is a general term used to describe a clinical syndrome with multiple aetiology, in which there is a progressive decline in areas of function such as memory, reasoning, communication skills and the ability to carry out daily activities (Department of Health, 2009; Knapp & Prince 2007). Alzheimer's disease (AD), the predominant symptom of which is memory loss, is the most common cause of dementia (Ferri et al., 2005). There is currently no cure for AD and only limited progress has been made with research into drug treatments (Schneider et al., 2014). Researchers have recognised that it may be more effective to target interventions at people in a pre-dementia phase of AD, before the progressive disease is established (Petersen et al., 2014).

Several terms, including age-associated memory impairment (AAMI) (Crook et al., 1986), aging-associated cognitive decline (AACD) (Levy, 1994) or cognitively impaired, not demented (CIND) (Ebly, Hogan, & Parhad, 1995; Ebly, Parhad, Hogan, & Fung, 1994) have been proposed in order to define the mild impairments that can precede established dementia. However, a concept which has become increasingly popular in clinical research and practice is the term Mild Cognitive Impairment (MCI). MCI is defined as "cognitive decline greater than that expected for an individual's age and education level but that does not interfere notably with activities of daily life" (Gauthier et al., 2006).

Concept of Mild Cognitive Impairment (MCI)

Mild cognitive impairment (MCI) was first introduced as a term to describe the intermediate stage between healthy ageing and dementia over 25 years ago (Reisberg et al., 1989). Reisberg and colleagues used the term to describe people who scored a rating of 3 on the Global Deterioration Scale (GDS). This scale (1-7) was developed for the assessment of primary degenerative dementia, with stages 1-3 describing pre-dementia phases (where 1 = no cognitive impairment; 2 = subjective cognitive impairment; 3 = MCI) and stages 4-7 describing established dementia of increasing severity (Reisberg, Ferris, de Leon, & Crook, 1982). A Clinical Dementia Rating (CDR) scale score of 0.5 has also been applied to identify individuals with MCI. This is a scale used to stage the severity of AD and is derived from a semi-structured interview with the patient and a reliable informant (e.g. spouse or other family member/caregiver). Impairment in each of six cognitive categories (Memory, Orientation, Judgement and Problem Solving, Community Affairs, Home and Hobbies, and Personal Care) is rated on a five-point scale (where 0 = none, 0.5 = questionable, 1 = mild, 2 = moderate and 3 = severe) (Morris, 1993).

Whilst both the GDS and CDR are useful scales for staging the severity of dementia, they do not necessarily correspond to specific diagnoses (Petersen & Negash, 2008). In 1999, Petersen and colleagues further developed the concept of MCI by proposing definitive criteria to characterise individuals at high risk of further cognitive decline (Petersen et al., 1999). The original criteria (see Table 1.1) were specifically designed to characterize the early stages of an Alzheimer's like process, and were thus centred on memory impairment, consequently described as amnestic MCI (aMCI) (Petersen & Negash, 2008).

Table 1.1: Petersen's criteria for amnestic mild cognitive impairment

	Criteria
1.	Memory complaint, preferably corroborated by an informant
2.	Objective memory impairment for age
3.	Essentially preserved general cognitive function
4.	Largely intact activities of daily living
5.	Not demented

Subsequent work indicated that aMCI may not encompass all of the prodromal states of dementia and that a broader conceptualisation of MCI was necessary. In 2003, a conference of international experts on MCI was convened and more expansive criteria for MCI were proposed, which included the consideration of multiple types of cognitive impairment (Petersen, 2004; Winblad et al., 2004). A diagnostic flowchart was devised (see Figure 1.1) to differentiate between MCI with memory impairment (amnestic MCI, aMCI) and MCI with non-memory cognitive domain impairment (non-amnestic MCI, naMCI). These two subtypes were further divided into single and multiple domain types depending upon whether only one or more cognitive domains were impaired respectively.

An outcome framework, detailing the presumed outcome for each MCI subtype according to its presumed aetiology was also developed by Petersen and colleagues (2004) (see Figure 1.2). As the figure depicts, amnestic MCI subtypes with presumed degenerative aetiology are likely to represent a prodromal form of Alzheimer's disease. Whereas the non-amnestic subtypes that emphasize impairments in non-memory domains have a higher likelihood of progressing to non-Alzheimer's dementias such as frontotemporal dementia and dementia with Lewy bodies (Petersen, 2003; Petersen & Negash, 2008).

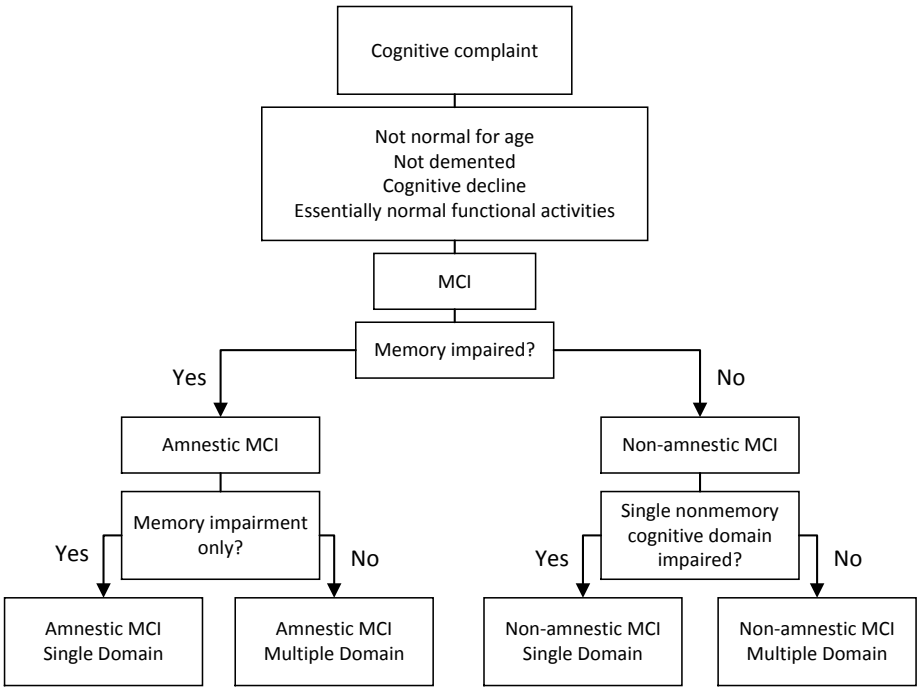


Figure 1.1: Flow chart of decision process for making diagnosis of subtypes of mild cognitive impairment. Adapted from Petersen (2004); Winblad et al. (2004)

		Etiology			
		Degenerative	Vascular	Psychiatric	Medical Conditions
Amnestic MCI	Single domain	AD		Depr	
	Multiple domain	AD	VaD	Depr	
Non-amnestic MCI	Single domain	FTD			
	Multiple domain	DLB	VaD		

Figure 1.2: Predicted outcome of MCI subtypes according to presumed etiology

Adapted from Petersen and Negash (2008)

(KEY: MCI = mild cognitive impairment; AD = Alzheimer's disease; FTD = frontotemporal dementia; DLB = dementia with Lewy bodies; VaD = vascular dementia; Depr = depression)

Since the publication of these expanded Petersen, or Mayo Clinic, criteria in 2003, numerous studies have used the definition to study the construct of MCI and they have become the most widely used criteria in the field. However, this is a continually developing field of research and recent advances in neuroimaging and neuropathology have led to the development of proposed biomarkers that reflect the neuropathological changes that are deemed to be the hallmark of AD, such as markers of Aβ deposition and of neuronal injury (Molin & Rockwood, 2016). These advances have led to the development of further diagnostic criteria, such as those proposed by the International Working Group for New Research Criteria for the Diagnosis of AD (Dubois et al., 2007) and the National Institute on Aging-Alzheimer's Association (NIA-AA) (Albert et al., 2011), which incorporate the use of biomarkers. However, their core criteria largely overlap with the Petersen criteria (Petersen et al., 2014) and the proposed biomarkers are not yet validated for use in routine practice (Molin & Rockwood, 2016; Petersen et al., 2014). Thus the diagnostic criteria proposed by Petersen and colleagues (2003) remain the most widely and routinely applied criteria and hence were the focus of the following work.

The prodromal state of MCI has been demonstrated with regard to neuropathological substrates (e.g. neuritic plaques, neurofibrillary tangles, hippocampal atrophy) in numerous studies (Luck, Luppa, Briel, & Riedel-Heller, 2010). For instance, a study by Petersen and colleagues (2006) compared autopsy tissue from 15 individuals who died whilst their clinical diagnosis was aMCI to that from age-matched groups of people who were either clinically healthy or had probable AD at the time of their death. Their study found that the aMCI individuals exhibited neuropathologic features which were intermediate between the neurofibrillary changes of aging and the pathologic features of early AD. In particular, all the patients with aMCI had pathologic findings involving

medial temporal lobe structures that likely accounted for their memory impairment, as well as many concomitant pathologic abnormalities, including argyrophilic grain disease, hippocampal sclerosis, and vascular lesions (Petersen et al., 2006). Another study by Grundman and colleagues (2004) found that people with aMCI had hippocampal volumes that were intermediate between those of controls and patients with AD (Grundman et al., 2004). Taken together this provides further evidence to suggest that MCI represents a transitional state between normal aging and dementia.

Incidence, Prevalence & Progression of MCI

A systematic review of population-based studies conducted by Luck and colleagues (2010) found that the reported incidence rates of MCI vary substantially, with rates ranging from 8.5 to 76.8 per 1000 person-years in persons aged ≥60 years (Luck et al., 2010). With regards to different subtypes, reports of aMCI incidence ranged from 9.9 to 40.6 per 1000 persons-years, and of naMCI incidence ranged from 28 to 36.3 per 1000 person-years. Five of the nine included studies reviewed the possible risk factors for incident MCI and found that higher age, lower education and hypertension were particularly associated with a higher risk of developing MCI (Luck et al., 2010).

A more recent systematic review by Ward and colleagues (2012) also found substantial variation in the reported prevalence rates of MCI, with rates of MCI ranging from 3 to 42% and of the aMCI subtype ranging from 0.5 – 31.9% (Ward, Arrighi, Michels, & Cedarbaum, 2012). This variation in both incidence and prevalence rates could be explained by the fact that there is currently no consensus on how MCI criteria should be operationalised, which has led to great variety in the methods employed to assess for MCI (Petersen et al., 2014). For instance, a wide variety of different procedures and tests have been used to assess cognition, as well as different cut-off points on test scores to define impairment (Ward et al., 2012). Other reasons for the wide disparity in

incidence and prevalence rates have been suggested, such as differences in sample size, recruitment strategies, geographic region and length of follow-up (Luck et al., 2010; Ward et al., 2012). The Medical Research Council (MRC) Cognitive Function and Aging Study (CFAS), a large-scale multi-centre study that was set up to investigate dementia and cognitive decline in a UK-based representative sample of more than 18,000 people aged over 65 years, conducted a direct comparison of different classifications of MCI in a subsample of their population to further investigate reported disparities in prevalence rates. Across the different classification systems, prevalence varied widely from 0.1% - 42%, reflecting differences in the focus and content of each system. For aMCI and multiple domain MCI specifically classified according to the Petersen criteria, they reported a prevalence of 2.5% (1.7 – 3.6%) and 2.6% (1.8 - 3.5%) respectively (Stephan, Matthews, McKeith, Bond, & Brayne, 2007).

A number of studies have also investigated the rates of progression of people with MCI to dementia. In the paper in which they first proposed diagnostic criteria for aMCI, Petersen and colleagues (1999) reported conversion rates to AD of 12% per year in a cohort of 76 people with aMCI who were followed up for 4 years. This was in contrast with a 1-2% conversion rate observed in the control subjects (Petersen et al., 1999). Other studies have reported similar rates, for instance, a study by Lonie and colleagues (2010) reported an annual conversion rate of 11.4% (95% CI 4–23%) to dementia (most often AD) in participants with aMCI (Lonie et al., 2010). The MRC CFAS study (referred to above) reported a 7.4% conversion rate of MCI to dementia at 2 year follow-up in the same subsample used to investigate prevalence rates. Again conversion rates varied depending on the classification system used, from 0.3 – 29.0% (Matthews, Stephan, McKeith, Bond, & Brayne, 2008). Other studies have reported even higher conversion rates. For instance a study by Geslani and colleagues (2005) reported a

conversion rate to AD of 41% after one year, based on a longitudinal follow-up of 54 aMCI participants (Geslani, Tierney, Herrmann, & Szalai, 2005). However, other studies with longer follow-up periods have reported lower conversion rates. A systematic review of 15 studies that had a follow-up period of 5 years or longer reported a pooled annual conversion rate to dementia of 4.2% (95% CI 3.9% to 4.6%) (Mitchell & Shiri-Feshki, 2008). It was observed by Mitchell and Shiri-Feshki (2008) that the highest rates of conversion were often seen in clinical samples recruited from specialist centres, such as memory clinics. The review also reported that the proportion converting to dementia tended to decline with longer observation periods, suggesting that the risk of progression diminishes with time. It was argued that this was likely to be due to the fact that in the first few years of follow-up many of those with the most adverse risk profile will tend to progress, dropout or die, leaving a cohort of less vulnerable sufferers. The authors pointed out, however, that an inverse temporal relationship was also evident in those who completed long term follow-up, suggesting other factors could be involved such as sampling issues or heterogeneity in the construct of MCI itself (Mitchell & Shiri-Feshki, 2008).

The instability of MCI has been highlighted in some studies. For instance, in the study by Lonie and colleagues (2010), of the aMCI participants who did not convert to dementia, 31% in fact had reverted back to normal cognition within 4 years. Other population-based studies have estimated that up to 44% of patients who have MCI at baseline return to normal cognitive function a year later (Ganguli, Dodge, Shen, & DeKosky, 2004; Ritchie, 2004). These studies underline the fact that there are many factors affecting cognitive performance in elderly populations apart from neurodegenerative disorders, including education, depression, anxiety, medication, vascular disease, and other treatable conditions, which could account for why many

cases of MCI are reversible (Petersen et al., 2014). Petersen argues that instability of the construct is perhaps not surprising due to the subtle nature of early impairment and it has been demonstrated that stability of the diagnosis improves with longer follow-up periods (Petersen et al., 2014).

Rationale for Detecting MCI

Despite the controversy surrounding the concept and diagnosis of MCI, it is clear that these individuals have a higher risk for developing neurodegenerative dementia, and are therefore of research and clinical interest. Dementia is one of the major causes of disability and dependency among older people worldwide (*Dementia: A Public Health Priority*, 2012). The burden of dementia is enormous, affecting the individual, as well as their family and friends, on personal, emotional, financial and social levels. A report to the Alzheimer's society in 2007 estimated that there were approximately 700,000 people living with dementia in the UK and that this was set to rise to 1.7 million by 2051 (Knapp & Prince, 2007). The same report estimated the financial cost of dementia to the UK to be £17.03 billion per annum, including formal care costs as well as the financial value of unpaid informal care provided by family and friends (Knapp & Prince, 2007).

Alzheimer's disease (AD) is the most common cause of dementia, accounting for 60-80% of cases (Barker et al., 2002). There is currently no cure for AD and clinical trials of interventions to meaningfully slow or prevent disease progression post-diagnosis have largely been unsuccessful (Schneider et al., 2014; Stephan & Brayne, 2014). Consequently, there have been calls to target intervention trials at individuals in a pre-dementia phase, possibly before irreversible neuronal damage has occurred. As discussed earlier, the concept of MCI has been developed to try to identify people who

are potentially in a transitional state between normal and pathological ageing and at a high risk of progressing to dementia and are thus important to consider for future early intervention trials.

In order to develop and test interventions, large cohorts of people with MCI need to be identified. However, the diagnosis of MCI, based on the Petersen criteria described earlier, involves in depth neuropsychological testing which is time consuming and burdensome. Shorter, less arduous cognitive measures would provide a more efficient and feasible method for researchers to identify people with probable MCI from large, community-based populations for participation in large scale, pragmatic intervention studies. Shorter procedures would also be of use to primary care clinicians who often lack the time, training and resources to perform in depth cognitive assessments (Artero & Ritchie, 2003). Quick tests, sensitive enough to detect MCI, could provide a practical approach to screening for people at high risk of developing dementia who may require referral to specialist services for further assessment and monitoring.

Aims of Research & Outline of Thesis

The overall aim of the current thesis was to investigate the effectiveness of a range of brief cognitive tests to be used for identifying people with MCI in an efficient and accurate manner. This aim was met by addressing four research questions which are addressed in the following Chapters.

Research Question 1: What brief cognitive tests have been used previously to identify people with aMCI and what is the evidence for their accuracy?

Research question 1 is addressed in Chapter 2, which covers the findings from a systematic review, providing a summary of the studies that have previously investigated

the validity of brief cognitive tests for identifying people with the amnestic form of MCI (aMCI). This Chapter details the reported validity and reliability results of previously investigated tests as well as provides a commentary on the methodological quality of these previous studies.

Research questions 2-4 are covered in Chapters 3-5, which describe the experimental work that has been carried out to investigate the validity of two brief memory tests identified from the literature (the Memory Alteration Test (M@T) (Rami, Molinuevo, Sanchez-Valle, Bosch, & Villar, 2007) and the Test Your Memory (TYM) test (Brown, Pengas, Dawson, Brown, & Clatworthy, 2009)) as well as reaction time task derived measures for detecting MCI.

Research Question 2: Can people with MCI be identified and recruited from the community in an efficient and timely manner?

Research question 2 is addressed in Chapter 3, which describes the recruitment of the cohort of older people upon whom the tests under investigation were validated. This Chapter details the assessment methods that were employed to classify people as having MCI or not. As discussed previously, the Petersen criteria provide a framework by which to determine whether or not a person has MCI; however there is currently no consensus on how these criteria should be operationalised (Petersen et al., 2014). For this study, a standardised protocol of neuropsychological tests was developed in order to assess for objective cognitive impairment. The tests included in the battery are widely used in clinical settings and have established normal reference values and cut-offs. A classification consistent with MCI was determined according to the Petersen criteria (Petersen, 2004).

Research Question 3: Are candidate brief cognitive tests (namely the Memory Alteration Test (M@T) and the Test Your Memory (TYM) test) accurate, reliable and usable in identifying people with aMCI from the community?

Research question 3 is addressed in Chapter 4, which reports on the findings from the validation work that was carried out to assess the effectiveness of the M@T and TYM for identifying people with aMCI from the recruited cohort. The aims of this work were threefold: to investigate (1) their sensitivity/specificity in detecting aMCI; (2) their test-retest reliability performance; and, (3) their clinical utility, assessed in terms of administration time and completion rates.

Research Question 4: Are reaction time (RT) derived measures accurate in identifying MCI?

Research question 4 is addressed in Chapter 5, which reports on an investigation into the utility of simple reaction time task derived measures for identifying cognitive impairment, as an alternative to the more traditionally used neuropsychological memory tests. Previous research has suggested that mean level of, and in particular variability in, processing speed performance is a key predictor of MCI/dementia (Christensen et al., 2005; Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000). The work in this chapter directly compares the ability of the M@T and TYM with these more objective measures for predicting cognitive status.

The thesis ends with a discussion of the main findings from each of the studies, placing them within the context of previous work and highlighting their implications. The limitations of the current work are discussed as well as proposals for future work.

Development Work: From Concept to Thesis

It is important to place this thesis within the context of the broader contributing development and grant work. The experimental work covered in Chapters 3 and 4 was funded by a National Institute for Health Research (NIHR) Research for Patient Benefit (RfPB) grant which was awarded in December 2011. I was a named co-applicant on the grant and contributed to the writing of the application and co-ordinated it's submission. In July 2012, I registered to complete a PhD using the NIHR funded work as its basis. As ideas for the PhD thesis developed, further complementary pieces of work including the systematic review of brief cognitive tests (Chapter 2) and investigation of reaction time task derived measures (Chapter 5) were conducted.

Chapter 2: A Systematic Review of the Diagnostic Test Accuracy of Brief Cognitive Tests to Detect Amnestic Mild Cognitive Impairment

Introduction

As discussed in Chapter 1, MCI is a term used to describe the transitional state between normal aging and established dementia (Petersen et al., 1999). Subtypes of MCI include amnestic (with memory impairment, aMCI) and non-amnestic (without memory impairment, naMCI). MCI of the amnestic-type is the commonest subtype, with a 2.6 fold increased incidence rate compared to naMCI being reported in a large longitudinal study by Roberts and colleagues (Roberts et al., 2012). People with MCI are at an increased risk of developing dementia and aMCI in particular is associated with elevated rates of conversion to Alzheimer's disease (Petersen et al., 2001), which is the most common form of dementia (Ferri et al., 2005). The reliable recognition of people with aMCI is therefore of particular interest to both researchers and clinicians.

It has been suggested that targeting interventions on people with aMCI might prevent or slow their decline into dementia (Petersen, Roberts, Knopman, Boeve, Geda, Ivnik, Smith, & Jack, 2009). However, the application of the Petersen criteria to diagnose aMCI is not straightforward and requires assessment by a trained specialist, along with considerable commitment from the patient to complete a complex battery of cognitive tests that are time consuming and can be fatiguing. Less demanding cognitive tests might have utility to provide a more efficient method to identify people with aMCI in research or clinical settings, but this would require test accuracy to have been confirmed.

This Chapter describes the findings from a systematic review that was conducted to identify the brief cognitive tests that have been used to identify people with aMCI and to evaluate the evidence for their accuracy. The work updates an earlier review (Lonie, Tierney, & Ebmeier, 2009) but is also more focussed, including only those tools that take less than 15 minutes to administer and incorporates a quality appraisal of the included studies, an aspect not covered in the previous review. In addition, new information on any predictive validity and reliability measures reported for the tools is included.

Methods

A systematic review was performed to describe the test accuracy of brief cognitive tests that have been used to identify people with aMCI. The methodology and reporting of this review followed standard guidance (Deeks, Bossuyt, & Gatsonis, 2010; Moher, Liberati, Tetzlaff, Altman, & Group, 2009).

Criteria for considering studies for this review

Prospective studies assessing the diagnostic test accuracy (DTA) of brief and simple cognitive tests used to identify people with aMCI (index tests) against a reference standard were considered for inclusion. Studies published in a language other than English were excluded. Only peer reviewed articles were included.

Participants

Participants were people with aMCI (single and multi-domain) diagnosed according to the Petersen criteria.

Index tests

Index tests considered for inclusion were those that were considered to be brief and simple cognitive tests, where "brief" was defined as: (1) taking less than 15 minutes to administer and "simple" was defined as: (2) not computer-based or requiring specialist equipment; and (3) not requiring specialist staff for administration. Studies assessing telephone administered or wholly carer/informant rated screening tools were excluded.

Reference standard

As it is the most widely used procedure for diagnosing aMCI, only those studies which used the Petersen criteria as the reference standard for verification of diagnosis were included in the review.

Search methods for identification of studies

The following databases were searched to identify studies for inclusion: MEDLINE, EMBASE, BIOSIS Previews, Web of Science, PsychINFO, LILACS, CINAHL, AMED, Cochrane Library, ASSIA, IBSS, PsychARTICLES, Scopus, Sociological Abstracts and ProQuest dissertations and theses. Databases were searched from 1999 – July 2013 (see Appendix 2.1 for the strategy used to search CINAHL).

Selection of studies

The search results were imported into Endnote, de-duplicated, and all titles and abstracts of the remaining citations were examined for potential eligibility. Following this initial "sifting" stage, full text copies of all potentially relevant articles were obtained and assessed for inclusion according to the stated eligibility criteria. Each study selection phase was performed by two independent reviewers and any disagreements in the selection of abstracts and full texts were settled by consensus.

Data extraction and management

Two independent reviewers extracted all data from included studies using a bespoke standardised data extraction form (see Appendix 2.2) and any disagreements within the extraction forms between reviewers were resolved by consensus. The extracted data included information on methods (e.g. study design, recruitment procedure) and participant characteristics. The operationalised reference criteria used and index tests, including cut-off points for diagnosis, were recorded. In addition, any validity and test-retest reliability measures reported for the index tests were also extracted. The measures for DTA included: sensitivity, specificity and AUC (area under the receiver operating characteristic (ROC) curve) for discriminating between aMCI and cognitively normal participants. Where sensitivity and/or specificity were reported for more than one cut-off, only the values reported as optimal by the author were extracted. In cases where no optimal value was stated, the cut-off providing the highest sensitivity was extracted. Where available, data concerning the ability of the test to predict future dementia were also extracted.

Assessment of methodological quality

Two independent reviewers assessed the methodological quality of the included studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool (Whiting et al., 2011), as recommended by The Cochrane Collaboration. QUADAS-2 involves a structured assessment using signalling questions in four domains: patient selection, index test, reference standard, and flow and timing (see Appendix 2.2). For each domain, a judgement of high, low or unclear risk of bias was made. A judgement of high risk was made if the answer to any of the signalling questions within a domain was “No”, a judgement of low risk was made if the answer to all of the signalling questions within a domain was “Yes”, and a judgement of unclear risk was made in all

other circumstances. Similarly, for overall assessment of risk of bias, a study was considered to be at high risk of bias if any of the domains were judged to be high, at low risk of bias if all of the domains were judged to be low and at unclear risk of bias in all other circumstances.

Statistical analysis and data synthesis

RevMan 5.2 software (http://tech.cochrane.org/revman) was used to construct 2 x 2 tables of index test performance (i.e. number of true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN)) using reported sensitivity and specificity values, as well as total number of participants and proportion of aMCI participants. These data were used to calculate the sensitivity and specificity with 95% confidence intervals (CI) and construct forest plots. Positive and negative predictive values (PPVs and NPVs) and positive and negative likelihood ratios (LR+ and LR-) were also calculated. Meta-analysis was performed, where appropriate, using STATA version 13.0 (StatCorp, 2013) software. Where sufficient studies were found ($n \geq 4$) that reported DTA data for the same test and cut-off, pooled estimates of sensitivity, specificity, LR+, LR- and summary diagnostic odds ratios were produced using a random effects bivariate model (Harbord, Deeks, Egger, Whiting, & Sterne, 2007; Reitsma et al., 2005) and heterogeneity was assessed using the I^2 statistic. Where four or more studies were assessing the same test at different cut-offs, summary ROC curves were produced using the STATA MIDAS module (Dwamena, Sylvester, & Carlos, 2010).

Results

Results of the search

The search identified 6431 citations, of which, 158 were considered as potentially relevant and the full articles obtained. Subsequently, 119 reports were excluded and 39 included (see Figure 2.1). Of these, 37 were cross-sectional DTA studies and two were longitudinal DTA studies that investigated the predictive validity of several index tests. There were 5766 aMCI and cognitively normal participants included in these studies. The mean prevalence of aMCI reported in the cross-sectional DTA studies was 42.4% but varied considerably from 3.1% to 72%. The majority of these studies (n = 32) recruited their aMCI population from secondary care settings, such as memory clinics and hospital departments, four studies recruited from the community and the remainder recruited from a mixture of secondary care and community based settings (n = 3). The characteristics of the included studies are summarised in Table 2.1.

In total, 42 brief and simple cognitive tests were investigated in the 39 studies. These tests are listed and described in Table 2.2. Thirty five of the index tests involve single tasks. The similar single task index tests have been grouped together, for example, four versions of clock drawing tests (CDTs), five verbal fluency tasks (VFTs) and three verbal learning tasks (VLTs), one of which has four scoring methods. The CDTs involve the participant drawing a clock and setting the time. The VFTs involve the participant naming as many words as they can within a certain time period (usually one minute), either from a certain category or beginning with a certain letter. The VLTs involve the participant recalling a word list after it has been read out to them. Seven of the index tests involve multiple tasks. By their nature, they tend to involve the assessment of more cognitive domains and take slightly longer to administer.

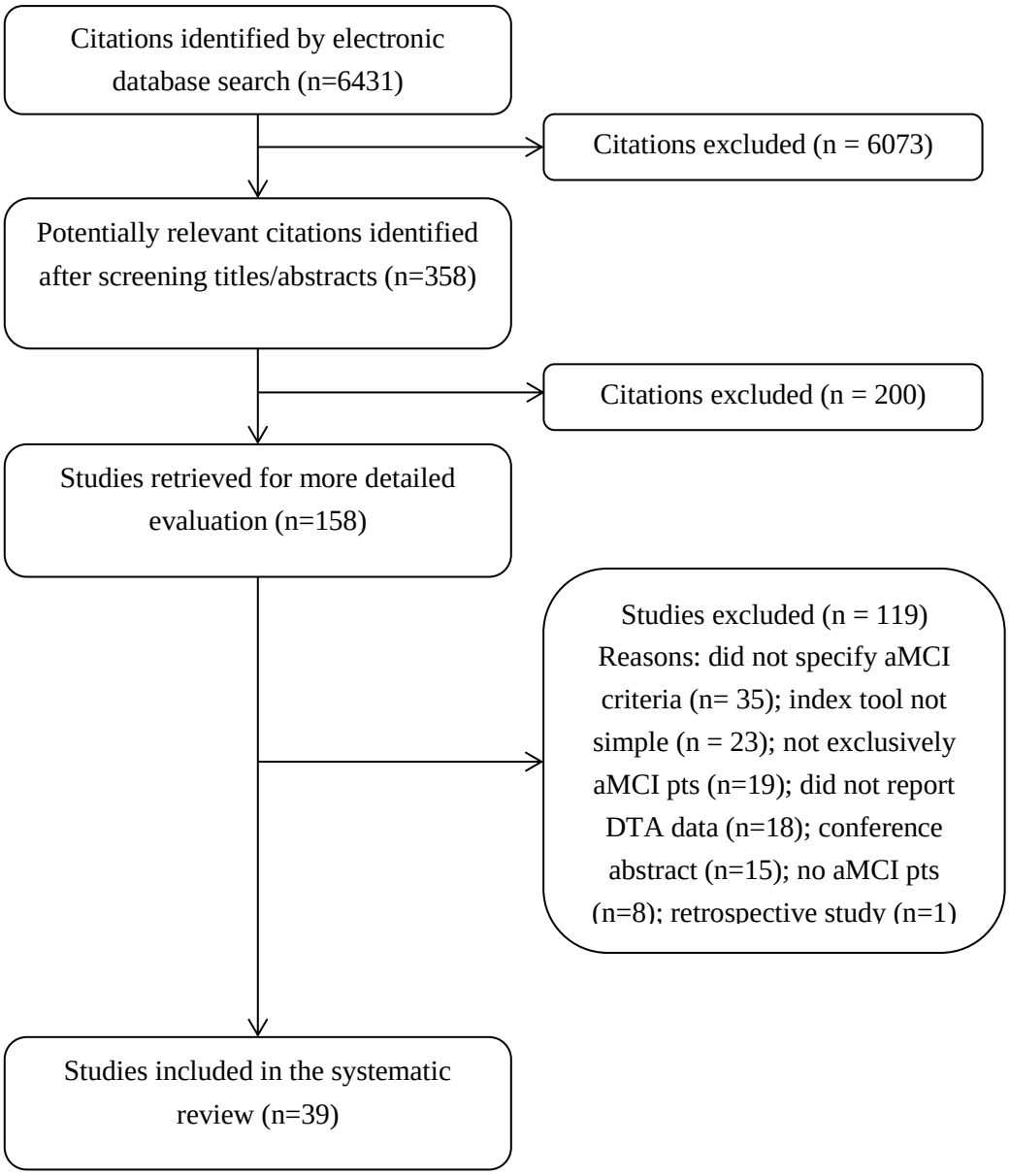


Figure 2.1: Study Selection Process Diagram (using the PRISMA guidelines)

Findings

DTA assessment

A summary of the DTA results reported for identifying people with aMCI is presented in Table 2.3 for single task index tests and Table 2.4 for multi-task index tests. Forest plots of sensitivity and specificity for each index test are presented in Appendices 2.3 and 2.4. Across the studies, sensitivity ranged widely from 7 to 100%, and specificity from 35 to 100%.

Table 2.1: Characteristics of included studies

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
Cross-sectional DTA studies								
Ahmed et al 2012	2012	UK	35	42.9	Community (OPTIMA cohort)	80.9 (7.2)	MoCA	Unclear
Ahn et al 2010	2010	Korea	120	35.8	Memory clinic	NR	MMSE	High
Alegret et al 2009	2009	Spain	88	50.0	Diagnostic unit of Fundacio ACE	76.5 (5.1)	15-Objects Test	High
Cacho et al 2010	2010	Spain	87	24.1	Memory clinic	73.8 (5.0)	CDT-command; MMSE (alone & in combination)	High
Chandler et al 2005	2005	USA	155	38.7	University Clinic for Alzheimer's and Related Diseases	72.8 (7.5)	VLT (CERAD-WLDR); MMSE	High
Costa et al 2012	2012	Germany	130	23.1	Memory clinic	67.8 (8.1)	MoCA	High
Dierckx et al 2007	2007	Belgium	92	43.5	Memory clinic & psychiatric hospital	75.0 (6.0)	MIS-plus & VAT	High
Diniz et al 2008	2008	Brazil	165	46.1	Memory clinic	72.3 (6.6)^{SD} 70.2 (6.5)^{MD}	MMSE	High
Freitas et al 2013	2013	Portugal	270	33.3	Dementia clinic	70.5 (8.0)	MMSE; MoCA	High
Fujiwara et al 2010	2010	Japan	66	45.5	Memory clinic	77.3 (6.3)	HDS-R; MoCA	High
Gonzalez-Palau et al 2013	2013	Spain	241	54.8	Memory clinic, residential facilities, community centres	82.0 (9.2)	VLT (HVLT LE); MMSE	High
Guo et al 2012	2012	China	508	61.2	Memory clinic	70.0 (9.1)^{SD} 70.3 (8.8)^{MD}	MES; MMSE	High
Hanyu et al 2009	2009	Japan	63	49.2	Memory clinic	75.6 (5.1)	VFT-Animals	High

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
Hanyu et al 2011	2011	Japan	80	57.5	Memory clinic	76.0 (6.6)	MMSE; TYM	High
Karrasch et al 2005	2005	Finland	30	50.0	Neurologist referral	67.5 (9.2)	CDT-CERAD; Constructional Praxis-Savings; Naming-BNT-M; VFT-Animals; VLTs (CERAD-WLDR, -WLLE, -WLRE, WLSA); MMSE	High
Kato et al 2013	2013	Japan	109	55.0	Hospital	76.1 (9.2)	CDT-command; MMSE	High
Ladeira et al 2009	2009	Brazil	166	50.0	Memory clinic	70.3 (6.1)	CDT-Sunderland; VFT-Animals; MMSE (alone & in combination)	Unclear
Lee et al 2008	2008	Korea	152	24.3	Hospital & community	71.3 (5.9)	MoCA	Unclear
Loewenstein et al 2009	2009	USA	103	22.3	Centre for AD and memory disorders	79.7 (6.0)	FBMS	High
Luis et al 2009	2009	USA	98	24.5	Memory clinic & community	78.9 (5.3)	MMSE; MoCA	Unclear
McLennan et al 2011	2011	Australia	98	3.1	Cardiac & diabetic/ endocrine outpatient clinics	NR	MoCA	Low
Muangpaisan et al 2010	2010	Thailand	107	72.0	Community (BLOSSOM cohort)	66.3 (7.9)	Digit Span (Forward & Backward); VFTs (Animals, Fruits, Letter Koh & Soh)	Unclear

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
Nasreddine et al 2005	2005	Canada	184	51.1	Memory clinic	75.2 (6.3)	MMSE; MoCA	High
Rahman et al 2009	2009	Egypt	184	51.1	Community (geriatric clubs)	NR	MoCA	Unclear
Rami et al 2007	2007	Spain	450	11.1	Memory-Alzheimer's Unit Hospital Clinic	76.6 (6.6)	M@T	High
Rami et al 2010	2010	Spain	87	57.5	Memory clinic	76.6 (6.6)	M@T	Unclear
Ravaglia et al 2005	2005	Italy	93	40.1	University Centre for Physiopathology of Aging	76.5 (7.1)	CDT (Sunderland, Wolf-Klein); MMSE (alone & in combination)	High
Saka et al 2006	2006	Turkey	51	35.3	Dementia outpatient clinic	69.4 (8.3)	ECR (3rd free & total recall)	High
Scheurich et al 2005	2005	Germany	20	65.0	Memory clinic	66.4 (9.7)	DemTect	Unclear
Schrijnemaekers et al 2006	2006	UK	73	26.0	Community (Foresight Challenge study)	76.2 (9.4)	MMSE; VLT (HVLT LE)	High
Smith et al 2007	2007	UK	35	65.7	Memory clinic	77.5 (7.8)	MMSE; MoCA	Unclear
Takahashi et al 2012	2012	Japan	50	50.0	Medical centre for dementia	75.2 (5.4)	AQT-CF	Unclear
Tsai et al 2012	2012	Taiwan	109	65.1	Memory clinic	79.2 (6.8)	MoCA	High
Woodard et al 2005	2005	USA	179	10.1	General Internal Medicine & Geriatric clinics	75.9 (5.7)	VFT-Animals; VLTs (CERAD-WLDR, -WLREDI, -WLSA)	Unclear
Yoshida et al 2012	2012	Japan	112	34.8	Memory clinic	71.4 (9.2)	MMSE	High
Zhao et al 2011	2011	China	300	50.0	Hospital	70.7 (4.3)	MoCA	High
Zhao et al 2012	2012	China	641^	50.7^	Memory clinic	74.1 (2.8)^	VLT (AVLT SR)	High
Longitudinal Studies								

Study	Year	Country	Total Number of Participants*	aMCI Prevalence (%)	aMCI Sample Source	aMCI Age (Mean (SD))	Index Test(s)	Risk of Bias
Ahmed et al 2008	2008	UK	18[A]	38.9[P]	Memory clinic	71.7 (6.8)[P] 71.3 (7.7)[NP]	Naming-GNT; TMT-Part B; VFT-Animals	High
Sarazin et al 2007	2007	France	217[A]	27.2[P]	Memory clinic	74.8 (4.1)[P] 70.9 (5.4)[NP]	FCSRT (Total & Free Recall); Serial digit ordering; Stroop - inhibition; TMT (A & B); VFTs (Fruits & "S"); WAIS (Similarities & Digit Symbol)	High

KEY: *aMCI and cognitively normal participants only; ^70-79yrs age group only; [A]aMCI participants only; [MD]multi-domain aMCI; [NP]aMCI non-progressors; [P]aMCI progressors; [SD]single domain aMCI; AD = Alzheimer's disease; NR = not reported

NB: for index test definitions see Abbreviations

Table 2.2: Description and cognitive domain coverage of the included index tests

Index Test	Brief Description	Max Score	TTA (min)	Memory	Semantic Knowledge/ Language	Visuospatial/ Perceptual processing	Attention / Orientation	Executive function / Fluency
Single Task Index Tests – MEMORY								
ECR	Name 16 items (following semantic cues) and recall them over 3 trials		7	Yes	-	-	-	-
-Free Recall	Number of free recall items per trial	16						
-Total Recall	Sum of free and cued recall items	48						
FBMS	Read and recall 15 presented words	15	3-4	Yes	-	-	-	-
FCSRT	Name 16 items (following semantic cues) and recall them over 3 trials		NR	Yes	-	-	-	-
-Free Recall	Sum of free recall items	48						
-Total Recall	Sum of free and cued recall items	48						
MIS-plus	Verbal cued recall task (6 words)	6	4	Yes	-	-	-	-
VAT	Name 6 pairs of objects/animals from line drawings, then recall 1 object after being presented with 1 from the pair	6	NR	Yes	-	-	-	-
VLT	Recall as many words as possible		NR	Yes	-	-	-	-
-AVLT SR	3 x 12 words, short delay recall	12						
-CERAD WLDR	3 x 10 words, short delay recall	10						
-CERAD WLLE	Sum of words recalled over 3 trials	30						
-CERAD WLRE	Recognition of words from list	100%						
-CERAD WLSA	Delayed recall adjusted for acquisition	100%						
-HVLT LE	3 x 12 words, sum of words recalled	36						
Construct. Praxis-Savings	Copy 4 geometric forms. Savings score is the proportion of elements remembered on delayed recall	100%	~5	Yes	-	Yes	-	-
Digit Span	Recall a series of numbers:	-	NR	Yes	-	-	Yes	-
-Forward	in the original order							
-Backward	in reverse order							

Index Test	Brief Description	Max Score	TTA (min)	Memory	Semantic Knowledge/ Language	Visuospatial/ Perceptual processing	Attention / Orientation	Executive function / Fluency
Serial Digit Ordering Test	Repeat a series of 7 digits, reordering them in ascending order for 15 trials	105	NR	Yes	-	-	Yes	-
WAIS-Digit Symbol	After being shown digit-symbol pairs, write down the corresponding symbol under each digit (time limited)	NR	1.5	Yes	-	Yes	Yes	-
Single Task Index Tests – NON-MEMORY								
Naming	Name objects from line drawings		NR	-	Yes	-	-	-
-BNT-M		15						
-GNT		30						
VFT	Name as many items as possible:	-	1	-	-	-	-	Yes
-Animals/Fruits	from a category							
-Koh/Soh/S	beginning with a certain letter (Koh)							
15-OT	Name all objects shown in a line drawing of 15 overlapping objects.	15	NR	-	Yes	Yes	Yes	Yes
AQT-CF	Timed test of naming colours & forms	Timed	3-5	-		Yes	Yes	-
CDT	Clock drawing (various versions)		1-2	-	Yes	Yes	-	-
-CERAD	No detail given	NR						
-Command	Draw clock and set time (11.10)	10						
-Sunderland	Draw clock and set time (2.45)	10						
-Wolf Klein	Scored based on number placement	10						
Stroop Test-inhibition	Name colours when the colour word and ink are incongruent	100	NR	-	Yes	-	Yes	-
TMT	Trace a line joining:	Timed	-	-	-	-	Yes	Yes
-A	numbers up in order							
-B	numbers and letters up in order							
WAIS-Similarities	Explain how two items might be similar	33	NR	-	Yes	-	-	Yes

Index Test	Brief Description	Max Score	TTA (min)	Memory	Semantic Knowledge/ Language	Visuospatial/ Perceptual processing	Attention / Orientation	Executive function / Fluency
Multi-Task Index Tests								
DemTect	5 tasks: a 10-word list learning task & delayed recall, number transcoding, semantic fluency, digit span reverse	18	8-10	Yes	-	-	Yes	Yes
HDS-R	9 questions/tasks: age, orientation in time & place, repeat & recall of 3 words, serial subtraction, digits backwards, recall of 5 objects, category fluency	30	NR	Yes	-	-	Yes	Yes
M@T	5 sections: Encoding & Free & Cued Recall (5 words & 2 sentences); Temporal Orientation; Semantic Memory	50	~5	Yes	Yes	-	Yes	-
MES	7 tasks: sentence repeat and short & long delay recall, category fluency, conflicting instructions, action imitation, inhibitory control test	100	~7	Yes	-	-	-	Yes
MMSE	6 sections: Orientation (time & place); Registration & Recall (3 words); Attention (serial subtraction/backward spelling); Language (naming, repeating, command following, sentence writing); Copying (intersecting pentagons)	30	~10	Yes	Yes	Yes	Yes	-
MoCA	8 sections: Visuospatial/Executive (TMT B, copy cube, CDT); Naming (animals); Registration & Recall (5 words); Attention (digit span, target detection, serial subtraction); Language (sentence repeat, letter fluency); Abstraction (similarity); Orientation (time & place)	30	10-15	Yes	Yes	Yes	Yes	Yes
TYM	Self-completed questionnaire with 10 tasks: Orientation; Sentence copying & Recall; Semantic Knowledge; Calculation; Category Fluency; Similarities; Naming; 2 x Visuospatial tasks (forming a letter & CDT)	50	5-10	Yes	Yes	Yes	Yes	Yes

KEY: NR = not reported; TTA = time to administer

NB: for index test definitions see Abbreviations

Table 2.3: Summary of diagnostic test accuracy results for single task cognitive tests for identifying aMCI

Index Test (units)	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
15-Objects Test	12	64	86	82	70	4.67	0.42	0.85	(Alegret et al., 2009)
AQT-CF (seconds)	72/73	84	76	78	83	3.50	0.21	0.88	(Takahashi et al., 2012)
Clock Drawing Test-									
CERAD	5	7	87	33	48	0.50	1.08	NR	(Karrasch et al., 2005)
-Command	8/9	76	70	44	90	2.51	0.34	0.78	(Cacho et al., 2010)
		50	88	83	59	4.08	0.57	0.72	(Kato et al., 2013)
-Sunderland	ELD*	30	88	71	56	2.50	0.79	0.59	(Ladeira et al., 2009)
	≤5	26	85	56	63	1.81	0.86	NR	(Ravaglia et al., 2005)
-Wolf Klein	≤6	21	89	57	62	1.93	0.89	NR	(Ravaglia et al., 2005)
Constructional Praxis-Savings (%)	60	33	67	50	50	1.00	1.00	NR	(Karrasch et al., 2005)
Digit Span									
-Forward	12	64	70	84	43	2.12	0.52	0.71	(Muangpaisan et al., 2010)
-Backward	4	77	57	82	49	1.77	0.41	0.73	(Muangpaisan et al., 2010)
ECR									
-3rd Free Recall	9	56	79	59	76	2.62	0.56	0.69	(Saka et al., 2006)
-Total Recall	42	50	91	75	77	5.50	0.55	0.63	(Saka et al., 2006)
FBMS	7[OR]	83	88	66	95	6.61	0.20	0.90	(Loewenstein et al., 2009)
Naming-BNT-M	11	13	100	100	53	-	0.87	NR	(Karrasch et al., 2005)
Verbal Fluency Task									
-Animals	ELD[$]	27	95	85	56	5.50	0.77	0.61	(Ladeira et al., 2009)
	14	81	69	71	79	2.58	0.28	NR	(Hanyu et al., 2009)

Index Test (units)	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
		83	43	79	50	1.47	0.39	0.63	(Muangpaisan et al., 2010)
	15	27	100	100	57	-	0.73	NR	(Karrasch et al., 2005)
	<20	72	55	15	95	1.62	0.50	0.69	(Woodard et al., 2005)
-Fruits	15	68	63	83	43	1.84	0.51	0.69	(Muangpaisan et al., 2010)
-Letter Koh	9	50	73	83	36	1.88	0.68	0.66	(Muangpaisan et al., 2010)
-Letter Soh	7	81	57	83	53	1.86	0.34	0.71	(Muangpaisan et al., 2010)
Verbal Learning Task									
-AVLT SR	≤2[^]	97	73	79	95	3.55	0.05	0.94	(Zhao et al., 2012)
-CERAD WLDR	6	27	100	100	57	-	0.73	NR	(Karrasch et al., 2005)
	6.5	82	63	58	85	2.22	0.29	0.82	(Chandler et al., 2005)
	<7	83	60	19	97	2.10	0.28	0.76	(Woodard et al., 2005)
-CERAD WLLE	20[OS]	73	80	79	75	3.67	0.33	NR	(Karrasch et al., 2005)
-CERAD WLREDI	<10	94	35	14	98	1.45	0.16	0.73	(Woodard et al., 2005)
-CERAD WLRE (%)	92[OS]	47	93	88	64	7.00	0.57	NR	(Karrasch et al., 2005)
-CERAD WLSA (%)	<80	89	55	18	98	1.99	0.20	0.77	(Woodard et al., 2005)
	80	33	67	50	50	1.00	1.00	NR	(Karrasch et al., 2005)
-HVLT LE	≤15[OR]	83	65	74	76	2.39	0.26	0.84	(Gonzalez-Palau et al., 2013)
	24.5	84	80	59	93	4.13	0.20	NR	(Schrijnemaekers et al., 2006)

KEY: [^]70-79 years age group; AUC = Area Under Receiver Operating Characteristic (ROC) curve; ELD* = education-level dependent (0-8 years of education <6; >8 years of education <8); ELD[$] = education-level dependent (illiterate <10, 1+ year of education <14); LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio; NPV = Negative Predictive Value; NR = Not Reported; PPV = Positive Predictive Value; [OR]indicates >1 threshold was reported in the study but only author-reported optimal threshold was extracted; [OS]indicates >1 threshold was reported in the study but only threshold with maximum sensitivity was extracted

NB: for index test definitions see Abbreviations

Table 2.4: Summary of diagnostic test accuracy results for multi-task cognitive tests for identifying aMCI

Index Test	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
DemTect	≤13	85	86	92	75	5.92	0.18	0.92	(Scheurich et al., 2005)
HDS-R	28/29	87	61	65	85	2.23	0.22	0.86	(Fujiwara et al., 2010)
M@T	37[OR]	96	79	36	99	4.57	0.05	0.93	(Rami et al., 2007)
		96	70	81	93	3.23	0.06	0.88	(Rami et al., 2010)
MES	≤72	88	91	91	88	10.2	0.13	0.96	(Guo et al., 2012)[MD]
	≤75	79	83	73	87	4.60	0.25	0.89	(Guo et al., 2012)[SD]
MMSE	<24	29	87	61	64	2.27	0.81	NR	(Ravaglia et al., 2005)
	24v25	52	95	79	86	11.5	0.50	0.82	(Cacho et al., 2010)
	25	13	93	67	52	2.00	0.93	NR	(Karrasch et al., 2005)
	<26	18	100	100	54	-	0.82	NR	(Nasreddine et al., 2005)
	≤26[OR]	76	69	75	71	2.43	0.34	0.76	(Gonzalez-Palau et al., 2013)
	26	17	100	100	39	-	0.83	NR	(Smith et al., 2007)
	26/27	63	96	95	68	15.5	0.38	0.84	(Kato et al., 2013)
		41	99	96	76	41.0	0.60	NR	(Yoshida et al., 2012)
	≤27	68	61	50	77	1.77	0.52	0.67	(Guo et al., 2012)[SD]
		68	70	69	69	2.26	0.46	0.72	(Guo et al., 2012)[MD]
		58	84	54	86	3.60	0.50	0.76	(Luis et al., 2009)[OR]
	27/28	72	60	33	88	1.78	0.47	0.72	(Diniz et al., 2008)*
		71	61	51	78	1.80	0.48	0.73	(Diniz et al., 2008)[$]
		70	68	74	62	2.15	0.45	0.73	(Hanyu et al., 2011)
	28.5	60	70	53	76	2.02	0.56	0.72	(Ahn et al., 2010)
		67	61	52	74	1.71	0.55	0.69	(Chandler et al., 2005)
		74	69	45	88	2.34	0.38	NR	(Schrijnemaekers et al., 2006)
	<29	67	72	71	68	2.40	0.46	0.75	(Freitas et al., 2013)
	ELD	54	71	65	61	1.88	0.64	0.63	(Ladeira et al., 2009)

Index Test	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
MoCA	<22	81	77	78	80	3.48	0.25	0.86	(Freitas et al., 2013)
	22/23[OR]	89	84	65	96	5.70	0.13	0.94	(Lee et al., 2008)
	23[OR]	96	95	85	99	17.7	0.04	0.97	(Luis et al., 2009)
	23.5	88	65	67	87	2.50	0.19	0.89	(Ahmed et al., 2012)
	23/24	86	86	-	-	5.93	0.17	-	(Fujiwara et al., 2010; Lee et al., 2008; Tsai et al., 2012; Zhao et al., 2011)[&]
	<24[OR]	100	50	6	100	2.00	0.00	NR	(McLennan et al., 2011)
	25/26[OR]	93	89	88	94	8.40	0.08	0.95	(Fujiwara et al., 2010)
	<26	90	87	88	90	6.77	0.11	NR	(Nasreddine et al., 2005)
		93	86	87	92	6.41	0.09	NR	(Rahman & El Gaafary, 2009)
	≤26	93	62	42	97	2.46	0.11	0.85	(Costa et al., 2012)
	26	83	50	76	60	1.65	0.35	NR	(Smith et al., 2007)
TYM	44/45	76	74	80	69	2.87	0.33	0.86	(Hanyu et al., 2011)

KEY: [&] meta-analysis of 4 studies; [MD]multi-domain aMCI; [SD]single domain aMCI; AUC = Area Under Receiver Operating Characteristic (ROC) curve; ELD = education-level dependent (illiterate: <20, 1-4 years education: <25, 4-8 years education: <26, 9+ years education: <28); LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio; NPV = Negative Predictive Value; NR = Not Reported; PPV = Positive Predictive Value; [OR]indicates >1 threshold was reported in the study but only author-reported optimal threshold was extracted; [OS]indicates >1 threshold was reported in the study but only threshold with maximum sensitivity was extracted

NB: for index test definitions see Abbreviations

Of the single task index tests, the Auditory Verbal Learning Test- Short Delay Recall task (AVLT-SR) showed the highest sensitivity (97%) for detecting aMCI, with a high specificity also (73%). The high AUC value (0.94) confirms the high diagnostic accuracy of the test. The Consortium to Establish a Registry for Alzheimer's Disease-Word List Memory Test Recognition Discrimination Index task (CERAD-WLREDI) also showed high sensitivity at 94% but very low specificity (35%). The delayed recall task from the same Word List Memory Test (CERAD-WLDR) showed high sensitivity in two of the studies (82%-83%) but low in one study (27%) and the Hopkins Verbal Learning Test- Wordlist Learning task (HVLT-LE) showed high sensitivity across two studies (83-84%). The Florida Brief Memory Screen task (FBMS), which also involves word recall, also showed high sensitivity (83%) and specificity (88%). The evidence of accuracy for the VFTs was less certain, with sensitivities ranging from 27% to 83%. This wide discrepancy in results may be explained by the variation in formats and cut-off scores used.

There were five studies that investigated the CDT as another single task index test. Again, results were fairly inconsistent across the studies with sensitivities ranging between 7% and 76%. Again, this variation in results may be explained by the different cut-off scores, administration and scoring methods used.

Of the multi-task index tests, the Mini Mental State Examination Scale (MMSE) was the most frequently investigated (17 studies). Sensitivity was reported for a number of cut-off values. The highest sensitivity reported was 76% for a cut-off value of $\leq$26. However, sensitivities were generally lower for other cut-offs, with most studies reporting sensitivities between 13% and 68%. The next most frequently investigated multi-task index test was the Montreal Cognitive Assessment (MoCA) with results

reported in 13 studies. Again, sensitivity was reported for a number cut-offs and the highest sensitivity reported was 100% at a cut-off of <24 but specificity was low at 50%. Sensitivities ranged from 81% to 96% for other reported cut-off values. Four studies reported sensitivity and specificity for the same cut-off of 23/24 and thus were combined in a meta-analysis (see Appendix 2.5). A pooled sensitivity and specificity of 86% was calculated at this cut-off value across the four studies (see Table 2.4).

Summary ROC curves were produced to provide a visual summary of the DTA reported across all studies investigating the VFT-Animals, MMSE and MoCA (see Figure 2.2). These curves illustrate that MoCA generally performed with higher sensitivity/specificity across studies than the VFT-Animals and MMSE, with most points gathering towards the top left hand corner of the ROC space (see Figure 2.2C). The high AUC value for the MoCA sROC curve (0.92, 95%CI 0.89-0.94) also confirms the higher diagnostic accuracy of the test in comparison with the VFT-Animals and MMSE, which had AUC values of 0.75 (95%CI 0.71-0.79) and 0.73 (95%CI 0.69-0.77) respectively.

Sensitivity was fairly high for all other multi-task index tests (ranging from 76% to 96%). Of these, the Memory Alteration Test (M@T) showed the highest sensitivity (96%) and fairly high specificity (70 – 79%) (see Table 2.4). The high reported AUC values (0.88 - 0.93) confirm the high diagnostic accuracy of this test.

Four studies reported DTA for combinations of index tests (see Table 2.5 and Appendix 2.6). The combination of MMSE and CDT-command showed the highest sensitivity for aMCI (76%).

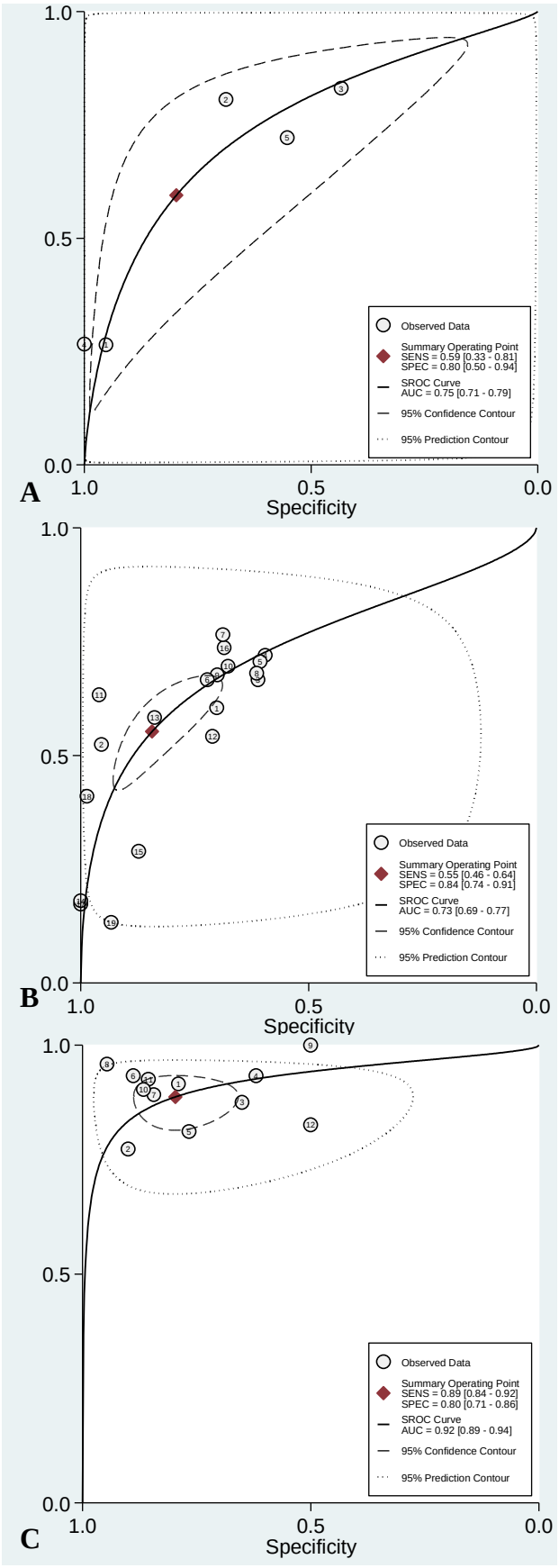


Figure 2.2: Summary receiver operating characteristic plots for studies using (A) VFT-Animals, (B) MMSE and (C) MoCA

Table 2.5: Summary of diagnostic test accuracy results for combined cognitive tests for identifying aMCI

Index Test (units)	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
MMSE									
& CDT-Command	35v36	76	77	52	91	3.35	0.31	0.86	(Cacho et al., 2010)
& CDT-Sunderland	ELD	19	98	89	55	8.00	0.83	0.58	(Ladeira et al., 2009)
& CDT-Sunderland & VFT-Animals	ELD	8	100	100	52	-	0.92	0.54	(Ladeira et al., 2009)
& VFT-Animals	ELD	17	100	100	55	-	0.83	0.58	(Ladeira et al., 2009)
OR CDT-Sunderland	<24 OR ≤5	45	69	50	64	1.45	0.80	NR	(Ravaglia et al., 2005)
OR CDT-Wolf Klein	<24 OR ≤6	37	71	47	62	1.27	0.89	NR	(Ravaglia et al., 2005)
OR CDT-Sunderland OR VFT-Animals	ELD	73	45	57	63	1.33	0.59	0.65	(Ladeira et al., 2009)
CDT-Sunderland & VFT-Animals	ELD	10	100	100	53	-	0.90	0.54	(Ladeira et al., 2009)
MIS-Plus & VAT	8	58	96	92	75	15.0	0.44	NR	(Dierckx et al., 2007)

KEY: AUC = Area Under Receiver Operating Characteristic (ROC) curve; ELD = education level dependent (CDT-Sunderland: 0-8 years of education <6; >8 years of education <8; VFT-Animals: illiterate <10, 1+ year of education <14; MMSE: illiterate <20, 1-4 years education <25, 4-8 years education <26, 9+ years education <28); LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio; NPV = Negative Predictive Value; NR = Not Reported; PPV = Positive Predictive Value

NB: for index test definitions see Abbreviations

Predictive Validity

Two longitudinal studies reported on the validity of 12 index tests for predicting future dementia over periods of one (Ahmed, Mitchell, Arnold, Nestor, et al., 2008) or three (Sarazin et al., 2007) years (see Table 2.6 and Appendix 2.7). The Free and Cued Selective Recall Reminding Test- Total Recall task (FCSRT-Total Recall) was the most accurate prognostic test with a sensitivity of 80% for identifying progressors and a specificity of 90% for identifying non-progressors. All other tests showed relatively low sensitivities.

Test-Retest Reliability

Test-retest reliability data were available for seven of the index tests (see Table 2.7). The MoCA was the most frequently investigated (eight studies). Most studies assessed reliability using the Intraclass Correlation Coefficient (ICC) with values ranging from 0.75-0.92 indicating a fairly high to a high reliability over a range of time periods from four weeks to 18 months. Reported ICCs for the MMSE tended to be slightly lower ranging from 0.67-0.76. High test-retest reliability was reported in one study for the A Quick Test of Cognitive Speed- colour-form naming task (AQT-CF) (ICC = 0.88). For other index tests, namely CDT, Florida Brief Memory Screen (FBMS) and MMSE & CDT, reliability data reporting was incomplete without a description of the method used to assess reliability (see Table 2.7).

Methodological quality

Of the 39 studies, 27 were assessed as high risk of bias; 11 were unclear risk of bias; and only one study (McLennan et al., 2011) scored a low risk of bias across all domains assessed (see Table 2.1 and Appendix 2.8). A summary of the quality assessment results across all four QUADAS-2 domains is provided in Figure 2.3.

Table 2.6: Summary of diagnostic test accuracy for single task cognitive tests for identifying aMCI-progressors (vs. non-progressors)

Index Test (units)	Cut-Off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	AUC	Study
FCSRT									
-Total Recall	40	80	90	75	92	7.87	0.23	0.94	(Sarazin et al., 2007)
-Free Recall	17	71	92	76	90	8.65	0.31	0.92	(Sarazin et al., 2007)
Naming-GNT	14	43	91	75	71	4.71	0.63	NR	(Ahmed, Mitchell, Arnold, Nestor, et al., 2008)
Serial Digit Ordering Test	80	58	68	40	81	1.79	0.63	0.77	(Sarazin et al., 2007)
Stroop Test-Inhibition	59	53	58	32	77	1.26	0.82	0.74	(Sarazin et al., 2007)
Trail Making Task									
-Part A (seconds)	53	63	59	36	81	1.52	0.63	0.73	(Sarazin et al., 2007)
-Part B (seconds)	128	50	55	44	60	1.10	0.92	NR	(Ahmed, Mitchell, Arnold, Nestor, et al., 2008)
	138	63	67	42	83	1.91	0.56	0.75	(Sarazin et al., 2007)
Verbal Fluency Task									
-Animals	11	29	100	100	69	-	0.71	NR	(Ahmed, Mitchell, Arnold, Nestor, et al., 2008)
-Fruits	13	56	82	54	83	3.16	0.54	0.80	(Sarazin et al., 2007)
-Letter “S”	17	58	56	33	78	1.32	0.75	0.74	(Sarazin et al., 2007)
WAIS									
-Similarities	11	49	72	40	79	1.77	0.70	0.78	(Sarazin et al., 2007)
-Digit Symbol Test	10	37	72	33	75	1.31	0.88	0.74	(Sarazin et al., 2007)

KEY: PPV = Positive Predictive Value; NPV = Negative Predictive Value; LR+ = Positive Likelihood Ratio; LR- = Negative Likelihood Ratio; AUC = Area Under Receiver Operating Characteristic (ROC) curve; NR = Not Reported

NB: for index test definitions see Abbreviations

Table 2.7: Test-retest reliability results

Index Test	Participants	Time Period	Measure	Result	Study
AQT-CF	Community, without dementia (n=22)	16 weeks	ICC	0.88	(Takahashi et al., 2012)
CDT	Cognitively normal (n=30) & Mild AD (n=30)	1-2 months	NR	0.98	(Cacho et al., 2010)
FBMS	MCI & AD (n=29)	8.9 (6.2) weeks	NR	r = 0.65	(Loewenstein et al., 2009)
MES	Cognitively normal, MCI & AD (n=30)	29.1 (5.8) days	Mean change	4.7 (5.8)	(Guo et al., 2012)
MMSE	Cognitively normal (n=30) & Mild AD (n=30)	1-2 months	NR	0.99	(Cacho et al., 2010)
	Cognitively normal (n=30)	3 months	ICC	0.76	(Freitas et al., 2013)
	Cognitively normal (n=30)	18 months	ICC	0.67	(Freitas et al., 2013)
	Cognitively normal (n=2), MCI (n=5) & AD (n=15)	4 weeks	ICC	0.76	(Yoshida et al., 2012)
MMSE & CDT	Cognitively normal (n=30) & Mild AD (n=30)	1-2 months	NR	0.99	(Cacho et al., 2010)
MoCA	Cognitively normal (n=10)	1 month	Paired t test	p=0.537	(Ahmed et al., 2012)
	Cognitively normal (n=30)	3 months	ICC	0.91	(Freitas et al., 2013)
	Cognitively normal (n=30)	18 months	ICC	0.88	(Freitas et al., 2013)
	NR	8 weeks	ICC	0.88	(Fujiwara et al., 2010)
	Status unknown (n=29)	4 weeks	ICC	0.75	(Lee et al., 2008)
	Cognitively normal, MCI & AD (n=26)	35 (17.6) days	ICC	0.92	(Nasreddine et al., 2005)
	Cognitively normal & MCI (n=26)	35 (17.6) days	Mean change	0.9 (2.5)	(Rahman & El Gaafary, 2009)
	Status unknown (n=20)	4 weeks	ICC	0.88	(Tsai et al., 2012)
	Cognitively normal (n=80) & MCI (n=80)	30 days	ICC	0.80	(Zhao et al., 2011)

KEY: ICC = Intraclass Correlation Coefficient; AD = Alzheimer's Disease; NR = Not Reported; MCI = Mild Cognitive Impairment

NB: for index test definitions see Abbreviations

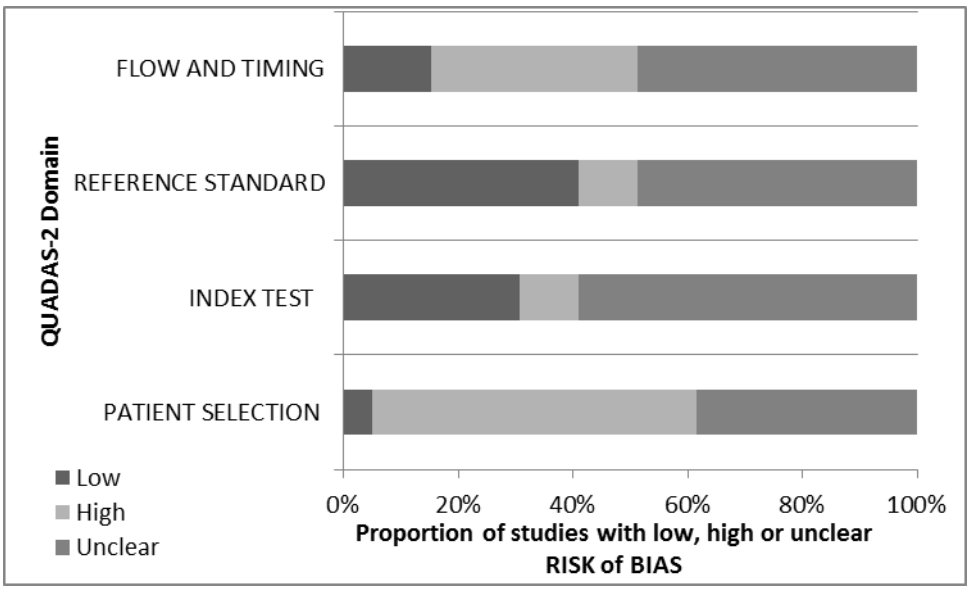


Figure 2.3: Summary of risk of bias judgements across all studies

Most studies (n=22) were judged to be at a high risk of bias in the patient selection domain due to unblinding of the participant assessment process resulting from the selection of people with known aMCI from memory clinics, and people with no cognitive impairment ("controls") from the community or from relatives of patients attending memory clinics. Twenty six studies were judged to be at an unclear risk of bias for the index or reference test interpretation since it was unclear the extent to which the tests were interpreted blindly. Ten of the studies were judged to be at high risk of bias in the flow and timing domain since patients and controls were not assessed with the same reference standard. Most studies (n=27) didn't report the time period between the index test and the reference standard and were therefore judged as unclear on this aspect.

Discussion

There is increasing interest in detecting people with aMCI as a potentially more timely point for treatment before the neuropathology has become more fully established with

consequent dementia. The practical difficulty is that the diagnostic criteria for aMCI (the Petersen criteria (Petersen, 2004)) are resource intense to apply in routine care. Brief cognitive tests have therefore been investigated as a more practical first step in providing a quick indication of a person's cognitive state. The idea is not that these brief tests would replace the standard diagnostic criteria but that they could be used to quickly identify people who may have aMCI and should be referred for further cognitive assessment. However, for them to be of use in identifying aMCI, a critical issue is to understand the diagnostic accuracy of the candidate tests. Therefore, a systematic search of the literature was conducted to find studies that had reported evidence on the DTA of brief cognitive tests for aMCI. To ensure their applicability for clinical settings, and for potential community screening for case ascertainment in research studies, only those tools characterised as simple (not requiring specialist input or equipment), and quick (less than 15 minutes to administer) were included. Evidence for 42 cognitive screening tools that met these criteria was found.

The AVLT-SR was the most accurate single task index test with a high sensitivity (97%), high specificity (73%) and high overall diagnostic accuracy (AUC = 0.94). Other verbal learning tasks (CERAD-WLDR and HVLT-LE), as well as FBMS, which involves word recall, also exhibited high sensitivity for aMCI (83-84%). The high accuracy of these word recall tests is perhaps unsurprising since they assess episodic memory, a feature known to be impaired in aMCI and early AD (Petersen et al., 1999) and thought to be the result of early pathological changes in the medial temporal lobe (Braak & Braak, 1998).

Although episodic memory impairment is an important distinguishing feature of people with aMCI, studies have shown that non-memory cognitive impairments such as

attention, processing speed, semantic fluency, executive function and visuospatial processing are also frequently detected in patients with aMCI (Economou, Papageorgiou, Karageorgiou, & Vassilopoulos, 2007). In this situation, the patient would be classified as multi-domain aMCI, that is, having cognitive impairment in memory and other non-memory domains. It has been reported that this form of aMCI may be more common than single-domain aMCI. For example, in a study by Alladi et al 2006, it was found that only 25 out of 90 patients with MCI had single-domain aMCI, and that deficits in both semantic memory and attention were more common (Alladi, Arnold, Mitchell, Nestor, & Hodges, 2006). Another study by Diniz et al 2008 reported a higher proportion of their patients had multi-domain aMCI compared with single domain (59% vs. 29% respectively) (Diniz et al., 2008) again supporting this idea. In addition, multi-domain aMCI may be more likely to progress to dementia than single domain aMCI (for review see (Hughes, Snitz, & Ganguli, 2011)).

For these reasons assessment for impairment in multiple cognitive domains might be important in the identification of people with aMCI. Some single task cognitive but non-memory tests have shown promising results in the reviewed literature. For example, the AQT-CF which assesses perceptual speed and attention has a sensitivity of 84%. Nonetheless, multi-task tests that assess several cognitive domains provide the potential for a more comprehensive assessment. Of the multi-task tests identified in this systematic review, all provide an assessment of memory but in combination with various other cognitive domains. The MMSE was the most frequently reported multi-task index test. However, the reported sensitivities were generally unsatisfactory in comparison to the other multi-task tests. The MoCA was the next most frequently reported multi-task test. This test provides an assessment of five cognitive domains and takes 10-15 minutes to administer. Although a sensitivity to detect aMCI of 100% has

been reported in association with a test score cut-off value of less than 24, for the most widely reported score cut-off value of 23/24 (four studies), the combined sensitivity was 86%. The M@T, which assesses episodic memory, semantic memory and orientation, and takes just 5-10 minutes to administer, also exhibited a very high sensitivity for aMCI (96%).

Another aim of this review was to identify evidence for the validity of brief cognitive tests for predicting future dementia in those with aMCI. Only two longitudinal studies were identified that investigated this issue. Of the 12 cognitive tests investigated, the FCSRT, which assesses free and cued item recall, had the highest sensitivity (80%) for identifying people with aMCI who progressed to dementia at three year follow-up. Clearly more longitudinal studies are needed to support these findings and to extend this aspect of validity to other cognitive tests.

Test-retest reliability is another important property of cognitive testing. Reliability data was reported for seven of the included brief cognitive tests. The reliability of the MoCA was reported in eight studies with fairly high to high reliability reported (ICC = 0.75-0.92). High test-retest reliability was also reported for the AQT-CF (ICC = 0.88), whereas the ICCs for the MMSE tended to be lower (0.67-0.76).

The methodological quality of the included studies was also assessed in this review and only one study (McLennan et al., 2011) had a “low risk of bias” in all four assessed domains. This study assessed the validity of MoCA for detecting aMCI patients recruited from hospital cardiovascular outpatient clinics and, although the MoCA detected all three patients with aMCI, it exhibited a low specificity of 50%. All other studies were judged to be at a high or unclear risk of bias and this therefore limits the

confidence with which interpretations from the studies can be made. A large proportion of the studies were judged to be at a high risk of bias in the patient selection domain since they were at risk of unblinding the patient assessment process by recruiting patients with known aMCI from memory clinics and participants without cognitive impairment ("controls") from the community or via relatives of the patients. It has been reported that studies such as these may exaggerate diagnostic accuracy (Lijmer et al., 1999; Whiting et al., 2004). Another area that was assessed as high risk of bias for several studies was patient flow, where studies did not use the same reference standard for all participants. Improved study design should be a feature of future studies in this area.

Strengths of the review

This review used systematic methods and followed standard guidance to provide a comprehensive summary of the literature on the diagnostic test accuracy of brief cognitive tests for aMCI. Two independent reviewers screened all potential studies for inclusion and extracted data, reducing potential risk of bias in study selection or errors in data extraction. All included studies were assessed for their methodological quality using a standardised tool (QUADAS-2). Finally, by ensuring that only those studies that used the Petersen criteria as the reference standard were included, the samples reported can be considered to be relatively homogeneous and comparable.

Weaknesses of the review

It is important to note that the mean prevalence of aMCI across the included studies was high at 42.4% (beyond the range of 0.5 – 31.9% reported in a systematic review of aMCI prevalence by Ward and colleagues (2012) and much greater than the rate of 14-18% for individuals aged 70 years and older estimated by Petersen and colleagues

(2009)). Therefore calculated estimates of PPV/NPV from these included studies are possibly inflated and unlikely to be generalizable to older people in community settings. Also, some studies reported multiple thresholds and, in these cases, the optimal threshold reported by the author was chosen. This may have led to an overestimation of diagnostic accuracy (Leeflang, Moons, Reitsma, & Zwinderman, 2008), particularly in the summary ROC curves, where two included studies for MMSE and five included studies for MoCA reported multiple thresholds.

Conclusion

An ideal cognitive test for detecting people with aMCI would be one with high parameter values for DTA, predictive validity and test-retest reliability in the context of a well-designed experimental study. Of the 42 brief cognitive tests identified in this review, the MoCA was identified as the most comprehensively investigated test. The MoCA has a high sensitivity and high test-retest reliability, but its predictive validity has yet to be investigated. Other brief cognitive tests, such as those that assess word recall (AVLT-SR, CERAD-WLDR, HVLT-LE and FBMS), and multi-task tests that assess several cognitive domains (such as M@T), have also been found to exhibit high sensitivities and reasonable specificities. However, lack of evidence on the predictive validity of these tests, and concerns over the quality of the constituent studies, limit the confidence with which definitive recommendations can be made. Further validation studies of the most promising cognitive tests to detect aMCI are warranted.

Chapter 3: The Identification of a Cohort of Older People with Mild Cognitive Impairment from the Community

Introduction

The identification of people at risk of developing dementia is becoming increasingly important with the prospect that early intervention might delay their progression to dementia. As discussed in Chapter 1, mild cognitive impairment (MCI) has emerged as a term to capture the pre-dementia phase of cognitive dysfunction (Petersen, Roberts, Knopman, Boeve, Geda, Ivnik, Smith, Jack Jr, et al., 2009) and is associated with elevated rates of progression to dementia (Petersen et al., 2001). The accurate detection of individuals with MCI has important implications regarding the development of dementia preventative interventions and future clinical trial recruitment (Stephan & Brayne, 2014).

Many brief cognitive tests have been developed for the detection of MCI, particularly the amnestic form of MCI, where memory impairment is the dominant symptom (see Chapter 2). However, these have largely been evaluated in the context of secondary care settings, such as memory clinics, with the performance of patients with known cognitive impairment being compared to unimpaired controls recruited from the community or via relatives of the patients. Studies which use such recruitment strategies are at high risk of unblinding the assessment process and may exaggerate diagnostic accuracy (Lijmer et al., 1999; Whiting et al., 2004). To address this limitation of previous validation studies, the current study investigated an alternative approach to recruitment which involved inviting volunteers from the community, without prior knowledge of their cognitive status, to form a cohort on which to validate some brief cognitive tests for identifying MCI. The following Chapters 4 and 5 report on the validity of the tests under

investigation. However, the focus of this Chapter is to report on the recruitment methods and assessment procedures that were employed to identify the cohort of older people.

The recruitment strategy was designed to reflect one that might be applied in the future, within a research context, to identify people with MCI for participation in dementia prevention trials. One of the most difficult challenges in clinical trials is whether appropriate participants can be identified and recruited in a timely manner and many trials either fail to reach recruitment targets or have to be extended (McDonald et al., 2006). This Chapter aims to provide valuable information on recruitment rates of people with MCI from a community-based cohort in order to provide an aid to planning future studies which seek to recruit older people with MCI.

Methods

Recruitment Process

The procedures involved in recruiting participants for the study are listed in Table 3.1. There were three screening stages involved in identifying potential participants for the study, followed by two steps to inform volunteers about the study and gain their informed written consent.

Screening Stage 1: GP Record Screening & Flyer Mail Out

The first stage in identifying potential participants involved GP practice managers screening their electronic medical records to search for people who met the following criteria: (1) were aged 70 years and older; (2) were not resident in a care or nursing home; (3) did not have dementia; (4) did not have current depression; (5) did not have

history of stroke within the previous three months; and (6) were not receiving palliative care. Study information flyers were then posted to all individuals who were found to be eligible from the electronic record screening. Recipients of the flyers were invited to answer some questions on the flyer and return it to our research office if they were interested in taking part in the study. A stamped addressed envelope was provided with the flyer to encourage responses. The GP practices involved were all based in Bradford, UK and had "research ready" status, accredited by the National Institute for Health Research (NIHR). Research Ready ® is a quality assurance programme intended for use by all research-active UK GP practices and designed in accord with the UK Research Governance Framework's legal, ethical, professional, and patient safety requirements ("Royal College of General Practitioners: RCGP Research Ready," 2016).

Screening Stage 2: Flyer Screening

The returned study information flyers were checked for eligibility by the research team. Over the study recruitment period (which ran from October 2012 to April 2015), the study information flyer and eligibility criteria were revised (see "Flyer Development" section for further details). According to the final version eligibility criteria, respondents were considered eligible to participate if they: (1) self-reported difficulty with their memory; (2) spoke English; (3) had attended school for at least eight years; and (4) had an informant available to answer some of the study questions.

The presence of a subjective memory complaint is included as a criterion within the Petersen framework for a MCI diagnosis (Petersen, 2004) and therefore only those self-reporting difficulty with their memory were primarily included in the study. However, there is some debate within the literature as to whether or not this criterion is essential for diagnosis (Mitchell, 2008b), and so it was decided that a sample of 100 people who

did not self-report memory difficulties should also be invited to take part. This enabled rates of MCI to be compared between those who did and did not report memory complaints.

The requirement to speak English was essential due to the application of cognitive tests which have not been validated in other languages and the requirement to have received at least 8 years of education was needed in order to fulfil the validity criteria of the IQ task used in the study.

Screening Stage 3: Telephone Screening

Those respondents who were eligible to take part (or required further clarification) were telephoned and invited to take part in the study. Further information on the study was provided at this stage and it was ascertained whether or not the person was medically stable and well enough to travel to our research offices. Appointments were subsequently made for a first home visit with those people who were eligible and still interested in taking part.

Final Steps 4 & 5: Informed Consent to Participate

Consent to participate in the study was obtained face-to-face during a visit to the participant's home. Participants were posted detailed study information sheets in advance of the appointment so that they had time to fully consider the study procedures before agreeing to take part.

Table 3.1: Recruitment procedures

Recruitment process	Description	Performed by
1. GP record screening and flyer mail out	GP electronic medical records screened to identify potentially eligible participants and subsequent flyers posted.	Practice manager
2. Flyer screening	Returned flyers checked for eligibility.	Research team
3. Telephone screening	Eligible respondents were telephoned to (i) outline the study, (ii) check further eligibility criteria and (iii) invite to participate. Session 1 appointments were made with eligible volunteers.	Research team
4. Participant and informant information sheets mailed	Information sheets mailed out prior to Session 1 visit.	Research team
5. Informed written consent sought	Participant and informant (if available) provide informed consent at Session 1 (in the participant's home) and participate in first assessments.	Research team

Flyer Development

The study information flyer was designed to include brief information about the study and some questions relating to eligibility criteria that could not be checked via the GP electronic record screening procedure. Clear and concise wording was used and drafts were checked by members of the Older People's Forum, which is a well-established group of consumer colleagues who meet regularly to discuss our Academic Unit's research, to ensure it was easily understood. During the study recruitment period, the study information flyer and eligibility criteria were revised (see Table 3.2).

Table 3.2: Flyer questions and eligibility criteria

	Flyer Version 1	**Flyer Version 2**
Memory Questions 1. Do you have any difficulty with your memory? 2. Do you forget where you left things more than you used to? 3. Do you forget the names of close friends or relatives? 4. Have you ever been in your own neighbour-hood and forgotten your way?	All four memory questions included.	Only memory question 1 included.
Mood Questions During the past month have you often been bothered by: 5. feeling down, depressed or hopeless? 6. having little interest or pleasure in doing things?	Both mood questions included.	Not included.
Eligibility Questions 7. Do you speak English? 8. Did you attend school for at least 8 years? 9. Do you have a husband/wife/relative/friend who would be willing to answer some questions during the study? 10. Are you well enough to travel to our research offices?	Questions 7-10 included.	Questions 7-9 included.
Eligible if	*Version 1.1* • Q1 or 2 or 3 or 4 = YES • Q5-6 = NO • Q7-10=YES *Version 1.2* • Q1 or 2 or 3 or 4 = YES, plus sample of Q1-4 = NO • Q7-9=YES	• Q1 = YES, plus sample of Q1 = NO • Q7-9=YES

The original flyer (Version 1) included ten questions in total: four questions from the Cambridge Mental Disorders of the Elderly Examination (CAMDEX) which are designed to identify people with subjective memory impairment (Roth et al., 1986); the two Whooley questions which are recommended by the National Institute of Health and Care Excellence (NICE) to screen for possible mood disturbance (NICE, 2009; Whooley, Avins, Miranda, & Browner, 1997); and four questions checking further eligibility criteria. According to the original eligibility criteria (Version 1.1), respondents were considered eligible if they (i) reported a memory complaint (i.e. answered yes to any of the four CAMDEX questions); (ii) screened negative for possible depression (i.e. answered no to both Whooley questions); (iii) answered yes to all four eligibility questions. Since depression has been shown to be associated with deficits in cognitive performance (Rock, Roiser, Riedel, & Blackwell, 2014), the aim was to exclude people with possible low mood from the outset to ensure that this factor did not confound our results.

However, upon inspection of the batch of flyers returned within the first month, it was discovered that large numbers of respondents were being excluded based on not meeting criteria (i) and (ii). A decision was therefore made to relax the eligibility criteria (from Version 1.1 to 1.2). It was decided that a sample of 100 people who did not report a memory complaint (i.e. answered no to all four CAMDEX questions) should be invited to take part and that people should not be excluded based on their answers to the two Whooley questions. Instead, a more in depth assessment of mood was conducted using the Geriatric Depression Scale (GDS) (Sheikh & Yesavage, 1986), once the participant had been enrolled onto the study. Twelve months later the flyer was simplified to remove redundant questions in order encourage a higher return rate (from version 1 to version 2, see Appendix 3.1 for the final flyer Version 2 that was used).

Ethics

The study was approved by the Yorkshire and The Humber National Research Ethics Service Committee (ref: 12/YH/0207). One of the major ethical issues for consideration in the design and set up of this study was whether or not to inform participants of the results of their cognitive assessments. Upon careful consideration, it was decided that, since the cognitive tests only provided indications of cognitive function and were not diagnostic or performed by clinicians, it would have been inappropriate to have informed participants of their results. This point was made clear to participants from the outset. In cases where the participant became particularly anxious or concerned during testing, they were given the option of discussing their worries with a clinician based at the Academic Unit where the research was taking place, or encouraged to discuss any further concerns with their GP. Participating GP practices were regularly sent a list of the names of their participating patients so that they were aware of who was taking part in the study.

Assessment Process

Participants were administered a standardised protocol of neuropsychological tests (see Table 3.3) at a second session which was scheduled to take place within two weeks of the participant consenting to take part in the study. The tests were administered by trained researchers and took place in a well lit room within the Clinical Research Facility based at the Bradford Royal Infirmary.

Table 3.3: Battery of cognitive and functional assessments used to classify mild cognitive impairment

Cognitive/ Functional Domain	Test(s)	Domain Impairment Definition
Memory	• CVLT-II – Short Delay Free Recall • CVLT-II – Long Delay Free Recall	Short Delay and Long Delay free recall >1.5sd below mean of published norms
Executive Functions	• Brixton Spatial Anticipation Test (BSAT) • Trail Making Test Part A (TMT A) • Trail Making Test Part B (TMT B)	Number of errors on BSAT $\geq 5^{th}$ percentile and time taken on TMT B $\geq 10^{th}$ percentile of published norms
Language	• Graded Naming Test (GNT) • Pyramids & Palm Trees Test (PPT)	GNT<12 and PPT≤48
Visuo-spatial Function	• Visual Object & Space Perception (Letter Identification, Object Decision, Dot Counting, Number Location) • Clock Drawing Test	Impaired scores on any two or more tests: Letter Identification (<17); Object Decision (<15); Dot Counting (<9); Number Location (<8); CDT (<4))
ADLs	• Bristol Activities of Daily Living Scale	Score ≥10

As discussed in Chapter 1, there is currently no consensus on how the Petersen criteria for MCI should be operationalised. A standardised battery of tests was developed for this study to objectively assess for cognitive impairment and the included tests were chosen based on their regular use within clinical settings and the availability of established normal reference values and cut-offs. The tests assessed cognitive performance across a number of domains, including:

1. ***Memory***, which was assessed using the California Verbal Learning Test (CVLT), 2nd Edition (Delis, Kramer, Kaplan, & Ober, 2000). The CVLT tests the ability to encode and recall words. It involves the administrator reading out a list of 16 words, each of which belongs to one of four categories (vegetables, transport, furniture and animals). The participant is asked to recall the words immediately and this is repeated for five trials. Following this encoding period, the participant is asked to recall the 16 words after a short delay (of a few minutes, during which a distractor list of words is read aloud) and also after a long delay (of approximately 20 minutes). For each participant, the number of words correctly recalled was recorded and the CVLT-II Comprehensive Scoring System software was used to convert scores to the number of standard deviations below published norms (Delis et al., 2000).
2. ***Executive function and attention***, which was assessed using the Brixton Spatial Anticipation Test (BSAT) (Burgess & Shallice, 1997) and Trail Making Test Parts A & B (TMT A & B) (Reitan & Wolfson, 1985). The BSAT primarily measures the ability to detect a rule, to follow it, and to switch to a new rule. During the BSAT, participants are presented with a booklet of pages, each of which contains an array of 10 circles, one of which is coloured blue. The position of the blue circle changes from one page to the next and the changes are governed by a series of simple rules that alter without warning. Participants are presented with one page of the booklet at

a time and asked to point to where they think the blue circle will be on the next page, based on the rule inferred from previous pages. For each participant, the total number of errors (i.e. incorrect position predictions) was recorded. The TMT is a test of visual attention and task switching. The task requires the participant to draw connecting lines between a sequence of consecutive circular targets on a sheet of paper. In Part A, the targets are all numbers (1-25) and the participant is instructed to connect them in sequential order. In Part B, the targets are numbers (1-12) and letters (A-L) and the participant is instructed to alternate between them. The participant is asked to complete the tasks as quickly and accurately as possible. The time taken to complete each part was timed using a standard stopwatch and recorded for each participant. Using published norms, these times were converted to percentiles (Tombaugh, 2004).

3. ***Visuospatial function***, which was assessed using the Visual Object and Space Perception (VOSP) battery (Warrington & James, 1991) and the Clock Drawing Test (CDT) (Reitan & Wolfson, 1985). In order to minimise testing time, only four of the eight VOSP subtests were applied in this study: two object perception subtests (Incomplete Letters and Object Decision) and two spatial perception subtests (Dot Counting and Number Location). In the Incomplete Letters subtest, the participant is shown 20 incomplete letters and asked to name or identify them. The Object Decision subtest involves the participant being shown pages on which silhouettes of four objects are presented. Only one of the silhouettes represents a real object and the participant is asked to point to the one that they think is real. In the Dot Counting task, participants are asked to count how many black dots there are on a page. For the Number Location task, participants are shown pages which present two squares arranged one above the other. The top square contains numbers arranged randomly and the bottom square contains one black dot. The participant is

asked to identify which number corresponds to the black dot. The number of correct answers on each subtest was recorded for each participant. The CDT is used as a measure of spatial dysfunction and neglect. The CDT can be performed in different ways and various scoring procedures have been proposed (Agrell & Dehlin, 1998). In this study the free-drawn method was used, where the participant was asked to draw a clock face, including all numbers, and set the time to "10 past 5". The clock was scored on five features (clock face, number quantity, number distribution, hour hand position, minute hand position) with a point awarded for each correct feature. The total score out of 5 was recorded for each participant.

4. ***Language***, which was assessed using the Graded Naming Test (GNT) (Warrington, 1997) and Pyramids and Palm Trees test (PPT) (Howard & Patterson, 1992). The GNT assesses object naming ability and requires the participant to name drawings of 30 objects ordered in ascending difficulty. The PPT assesses the capacity to access detailed semantic information about words, necessary for the identification of associations between two perceptually and functionally distinct entities. The PPT can be administered as a word and/or picture version; in this study, the word version only was applied. During the task, participants are presented with 52 triads of words (e.g. Pyramid, Palm Tree, Fir Tree). The target word (e.g. Pyramid) is always presented above the other two and, for each triad, the participant is asked to select which of the two bottom words (e.g. Palm Tree or Fir Tree) is semantically related to the top word. For both tasks, the total number of correct answers was recorded for each participant.

In addition, the National Adult Reading Test (NART) (Nelson & Willison, 1991) was administered to provide an indication of pre-morbid verbal IQ. Mood was also assessed, initially via the two depression screening questions included in the study information

flyer. Following their removal from the flyer, a more detailed assessment of mood was completed using the Geriatric Depression Scale-short form (GDS) (Sheikh & Yesavage, 1986), once the participant had been enrolled onto the study. The GDS was administered to the majority (93%) of participants (the remainder of the participants had been enrolled onto the study prior to the introduction of the GDS to the testing schedule).

Activities of daily living (ADL) performance was assessed using the Bristol Activities of Daily Living Scale (BADLS) (Bucks, Ashworth, Wilcock, & Siegfried, 1996). This is an informant-rated questionnaire consisting of statements to rate 20 daily living activities, designed specifically to reveal the everyday abilities of people who have memory difficulties. Each activity is rated on a four-point scale to indicate the level of independence with which the person can perform the task, ranging from a score of 0 which indicates that no help is required through to a score of 3 which indicates that the person is unable to complete the task even with supervision. This produces a total score range of 0–60. Participants were asked to attend the second session with a spouse, relative or friend who could answer these questions. The informant was asked to provide written informed consent prior to answering any questions.

Classification Process

A classification flowchart, based on the flowchart of the decision process for making diagnoses of MCI subtypes proposed by Petersen et al (Petersen, 2004), was devised to classify participants (see Figures 1.1 and 3.1). The cut off scores used to determine cognitive domain impairment are listed in Table 3.3. Memory impairment was defined using a cut-off score of 1.5 standard deviations (SD) below the mean of published

norms on short and long delay recall on the CVLT. Although no particular cut-off score is specified in the Petersen criteria, a cut-off score of 1.5 SD below norms is generally recommended based on the fact that in the original description of the MCI cohort followed by Petersen and colleagues (1999), the MCI group's mean memory performance was 1.5SD below age- and education-matched control subjects (Petersen, 2004; Petersen et al., 1999). For the remaining cognitive domain test scores, either cut-off scores recommended in the test manuals were applied or 5^{th} – 10^{th} percentile cut–off scores were applied since they approximately correspond with a cut-off score of 1.5SD below the normative mean on a normally distributed curve.

The classification categories were as follows:

1. ***Amnestic MCI (aMCI)*** – this included participants who demonstrated impairment in memory and no impairment in ADLs. This category included both single and multi-domain aMCI.
2. ***Non-amnestic MCI (naMCI)*** – this included participants who demonstrated impairment in non-memory cognitive domain(s) and no impairment in ADLs. This category included both single and multi-domain naMCI.
3. ***Cognitive difficulties beyond MCI (>MCI)*** – this included participants who demonstrated impairment in one or more cognitive domains and also impairment in ADLs.
4. ***Low Mood*** – this included participants who scored ≥ 6 on the GDS. This cut-off score was selected as it has been cited as the optimal cut-off for identifying depression in numerous studies (Wancata, Alexandrowicz, Marquart, Weiss, & Friedrich, 2006).

5. ***Control*** - this included participants who did not demonstrate cognitive impairment in any one cognitive domain and did not meet the criteria for aMCI, or the other possible study classifications listed above.

Classifications were initially determined by the trained researcher administering the tasks and were verified by the study neuropsychologist (Dr Krist Noonan).

Sample Size Target

Sample size estimates for diagnostic test accuracy studies are calculated based on the desired confidence interval for the adequate estimation of sensitivity/sensitivity (Hajian-Tilaki, 2014). Wilson's 'score' method (Wilson, 1927) was used to calculate 95% confidence intervals for an observed sensitivity/specificity of 90% or 95% for a range of sample sizes (see Table 3.4). It has been reported that this method is particularly well suited to calculating confidence intervals in situations where the proportion is large, as is optimally the case with measures of sensitivity and specificity and it has the added advantage of being relatively straight forward to calculate (Newcombe, 1998). Confidence intervals were calculated using an online calculator provided by PEDro, the Physiotherapy Evidence Database (Herbert, 2013).

As can be seen from Table 3.4, as sample size increases, the width of the confidence intervals decreases. The aim at the outset of the study was to recruit 200 people with aMCI in order to achieve a lower bound confidence interval of 85% for an observed sensitivity of 90%. It was proposed that this could be achieved by approaching 4800 people with the study information flyer, assuming a 14% eligible response rate (based on 70% of flyers being returned, with only 20% of returns reporting a subjective memory complaint (Jonker, Geerlings, & Schmand, 2000)) and a 30% rate of aMCI in

those reporting subjective memory impairment (Benito-Leon, Mitchell, Vega, & Bermejo-Pareja, 2010; Mitchell, 2008a).

Table 3.4: 95% Confidence intervals for observed sensitivity/specificity of 90% or 95% for a range of sample sizes

n	Observed Sensitivity/Specificity	95% Confidence Interval
60	90%	(79.9%, 95.3%)
	95%	(86.3%, 98.3%)
100	90%	(82.6%, 94.5%)
	95%	(88.8%, 97.9%)
200	90%	(85.1%, 93.4%)
	95%	(91.0%, 97.3%)

Data Analysis

Response, eligibility and recruitment rates were calculated as percentages of the total population that were contacted or responded. Rates of the subtypes of MCI were calculated as percentages of the total population that were assessed or contacted. Sensitivity and specificity of subjective memory complaint status in detecting aMCI and of the two Whooley questions in detecting low mood as measured by the GDS were also calculated. Between-group differences based on cognitive classification in participant characteristics (including age, years of education, NART IQ, GDS score and the neuropsychological test battery scores) were explored using the Kruskal-Wallis H test (since the data were not normally distributed). Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Difference in gender proportion between the classification groups was analysed using the Chi-square test of homogeneity. These analyses were performed using SPSS Statistics v22 (IBM).

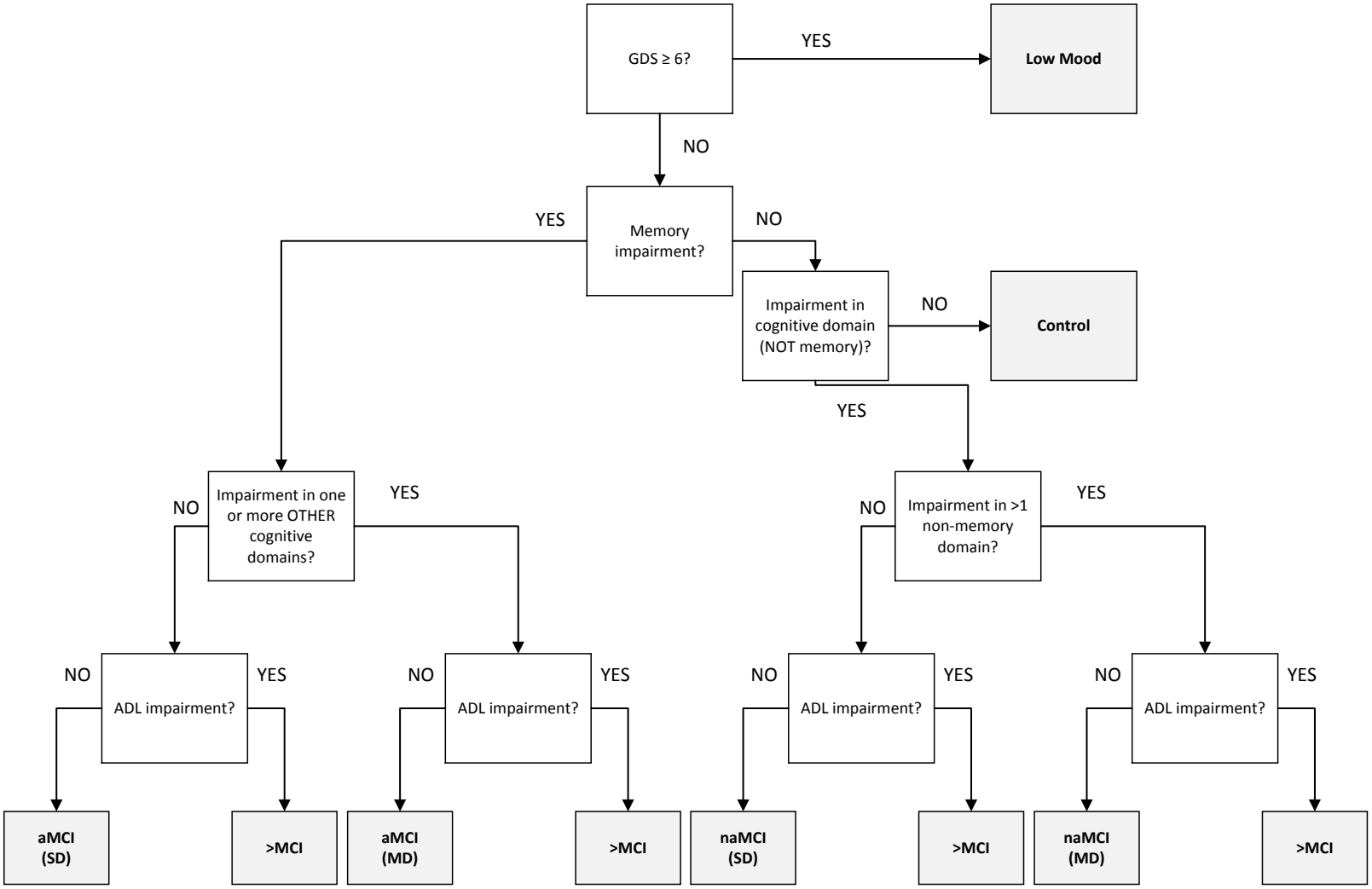


Figure 3.1: Classification flowchart

(adapted from (Petersen, 2004)). NB: GDS information not available for 34 participants (Whooley questions screen = negative)

Results

Figure 3.2 details the recruitment flow of participants throughout the study.

Screening Stage 1: GP Record Screening & Flyer Mail Out

A total of 85,870 medical records were screened across nine participating GP practices (average list size: 9541) and 7618 patients (8.9%) were found to be eligible and were sent screening flyers. The first 2600 patients were sent the original flyer Version 1 and, following revisions, the remaining 5018 patients were sent flyer Version 2.

Screening Stage 2: Flyer Screening

A total of 1477 flyers were returned, giving an overall return rate of 19.4%. The return rates from both flyer versions were similar (19.2% and 19.5% for flyer Version 1 and 2 respectively). The first batch of returned Version 1 flyers (n=126) were screened using the original eligibility criteria (Version 1.1, see Table 3.2) and only 19.8% (n=25) of this batch were deemed eligible. The most common reasons for being ineligible were having positive depression screening answers (54%) and no self-reported memory problems (40%). The remaining batch of returned Version 1 flyers (n=372) were screened using the revised eligibility criteria (Version 1.2, see Table 3.2) and the eligible rate increased to 73.4%. This time, the most common reason for being ineligible was having no informant available (90.7%). For the participants recruited from this later batch of flyers, the Whooley question depression screening data (Whooley et al., 1997) as well as GDS scores were available. The diagnostic test accuracy of the screening questions could therefore be estimated for the sample (see Table 3.5). It can be seen that the depression screening questions were 100% sensitive at

detecting people with low mood as measured by the GDS; however, they performed with lower specificity (71.8%) and 33 people who would have been excluded based on the original criteria were deemed not to have low mood upon further testing with the GDS.

Table 3.5: Diagnostic test accuracy of the Whooley depression screening questions

		GDS		Sensitivity (%)	Specificity (%)
		+	-		
Whooley screening questions	+	5	33	100	71.8
	-	0	84		

The remaining returned flyers were the revised version 2 flyers (n=979). Of these, 559 (57.1%) were deemed to be eligible and, again, the most common reason for being ineligible was having no informant available (n = 201, 49.5%). 174 (42.9%) people were excluded for having no self-reported memory problems (once the quota of 100 had been reached).

Screening Stage 3: Telephone Screening

In total, following the flyer screening stage, 857 people (58.0% of those people who returned flyers) were eligible for further follow-up by telephone. Of these, 557 (65.0%) were still interested in taking part and a home visit was arranged to consent them into the study. Those people who were excluded at this stage, included 157 (18.3%) who declined to take part, 30 (3.5%) who could not be contacted after numerous attempts and 113 (13.2%) who were ineligible. Again, the most common reason for being ineligible was having no informant available (n = 59, 52.2%).

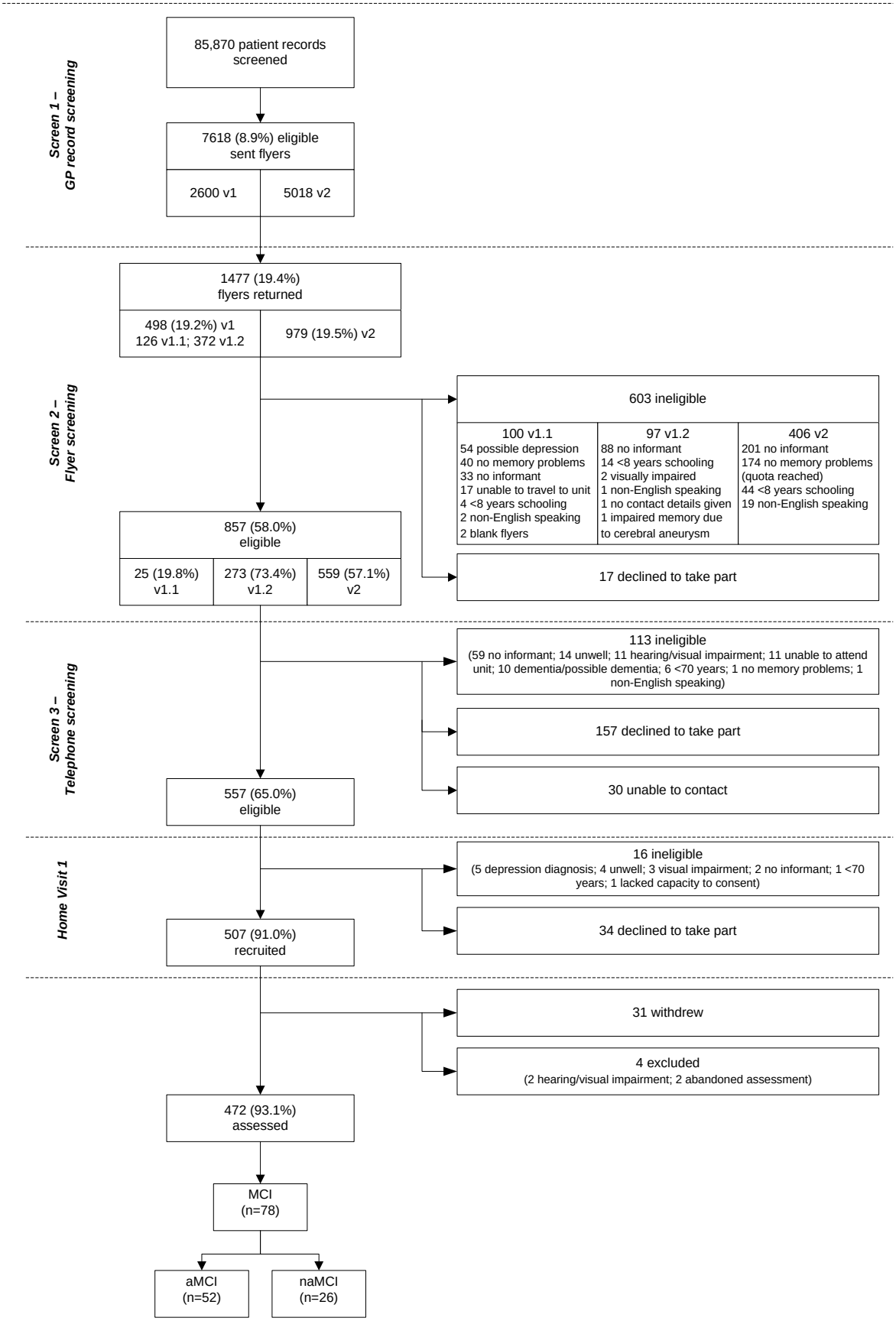


Figure 3.2: Recruitment flowchart

Recruitment & Assessment

The majority of those people who were eligible at the telephone screening stage went on to be recruited during a visit to their home (n = 507, 91.0%), and the majority of these people remained in the study and were assessed with the standardised battery of neuropsychological tests at a second session which took place at the BIHR and were assigned a cognitive classification (n = 472, 93.1%). Of those people not classified, 31 had withdrawn before the second session and four were excluded (two had visual/hearing impairment that affected their performance on the tasks and two could not complete the assessment due to distress or fatigue).

Classification Results

Figure 3.3 illustrates the proportions of the assessed participants that were grouped into each of the study classifications. Of the 472 people who were assessed, 78 people (16.5%) were classified as having some form of MCI. Of the people with MCI, 52 people (66.7%) were classified as having amnestic MCI (aMCI) and 26 people (33.3%) were classified as having non-amnestic MCI (naMCI). For both subtypes of MCI, single domain MCI was more prevalent than multi domain (n = 31, 59.6% for aMCI and n = 24, 92.3% for naMCI). Fourteen people were found to have cognitive difficulties beyond MCI (3.0% of the assessed population) and 20 people (4.2% of the assessed population) were classified as having low mood. The remaining 360 (76.3%) people did not demonstrate impairment in any one cognitive domain and were therefore classified as controls.

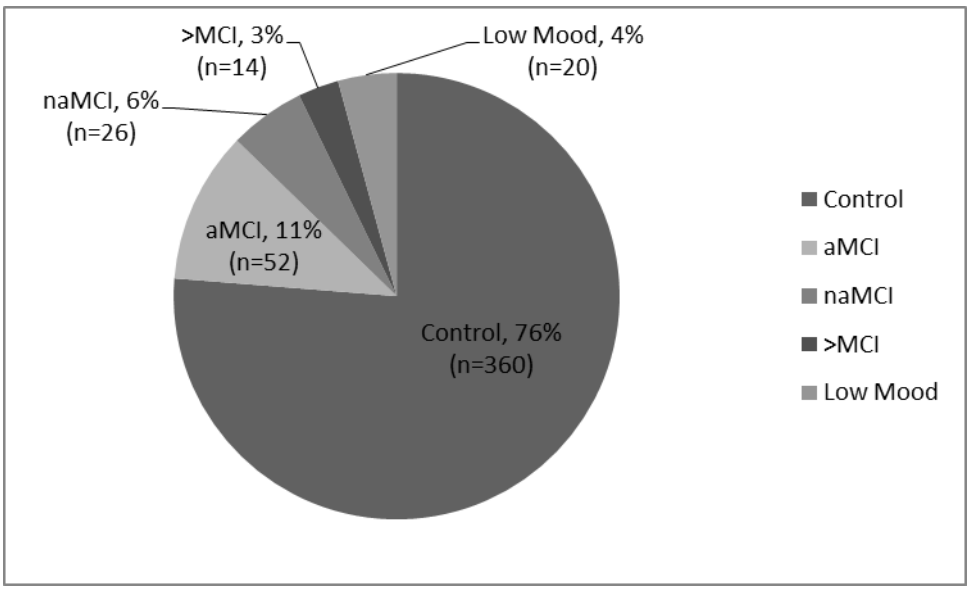


Figure 3.3: Cognitive classifications of assessed participants

KEY: aMCI = amnestic mild cognitive impairment; naMCI = non-amnestic mild cognitive impairment; >MCI = cognitive difficulties beyond mild cognitive impairment

Classifications by subjective memory complaint status

Table 3.6 shows the number of participants in each classification group with and without a memory complaint. Of those reporting a memory complaint, only 24% (n = 90) showed some form of cognitive impairment on formal testing (including 9 out of 19 people with low mood who demonstrated cognitive domain impairment), with 18.3% (n = 69) of people meeting the criteria for any form of MCI and 12.5% (n = 47) of people meeting the criteria for aMCI. 3.2% (n = 12) of those reporting a memory complaint had cognitive difficulties beyond MCI.

Of those reporting no memory problems, the majority (n = 84, 88.4%) did not demonstrate cognitive domain impairment on formal testing (including all controls and one participant with low mood). The remainder (n = 11, 11.6%) did show some impairment on cognitive testing, with 9.5% (n = 9) meeting the criteria for any form of MCI and 5.3% (n = 5) meeting the criteria for aMCI.

Table 3.7 reports how sensitive and specific subjective memory complaints were at identifying people with aMCI. As can be seen from the results, subjective memory complaints appear to be highly sensitive at picking up cases of aMCI (demonstrating sensitivity of 90%) but are not very specific (23%).

Table 3.6: Number of participants in each classification group with and without a memory complaint

Classification	Memory Complaint	No Memory Complaint
Control	277, 73.5%	83, 87.4%
aMCI	47, 12.5%	5, 5.3%
naMCI	22, 5.8%	4, 4.2%
>MCI	12, 3.2%	2, 2.1%
Low Mood	19 (9), 5.0% (2.4%)	1, 1.1%
Total	377	95

NB: Memory complaint defined as: answering yes to Q1 or 2 or 3 or 4 on flyer Version 1 (n=157) or answer yes to Q1 on flyer Version 2 (n=315); 19(9) indicates that 9 out of 19 people with low mood demonstrated cognitive domain impairment

Table 3.7: Ability of subjective memory complaint status to discriminate between aMCI and controls

	Total Sample*	aMCI participants	Control participants	Sensitivity (%)	Specificity (%)
Total, n	412	52	360	**90**	**23**
With SMC	324	47	277		
No SMC	88	5	83		

*excluding naMCI, >MCI and low mood participants

KEY: SMC = subjective memory complaint

Participant demographics and neuropsychological test scores

Demographic characteristics of the assessed participants are reported in Table 3.8. There were statistically significant differences between the groups in age ($\chi 2(4) =$

35.337, $p<0.001$), years of education ($\chi2(4) = 24.509$, $p<0.001$) and NART IQ ($\chi2(4) = 25.411$, $p<0.001$). Post-hoc pairwise comparisons revealed that the aMCI, naMCI and >MCI groups were significantly older than the controls, the aMCI and low mood groups had significantly fewer years of education than the controls and the low mood group had a significantly lower NART IQ than the controls.

Table 3.8 also reports the mean neuropsychological test scores of the assessed participants, categorised by their classifications (control, aMCI, naMCI, >MCI and low mood). Since the test scores formed the basis of the classifications, as expected, statistically significant differences were found between groups on all tests scores (at $p<0.02$). It is perhaps interesting to note that the low mood group performed significantly worse than controls on a number of cognitive tests, including CVLT short and long delay recall, Brixton errors, VOSP-Number Location and the Pyramids & Palm Trees test, indicating that their mood status may have had some impact on their cognitive performance.

Table 3.8: Participant characteristics and neuropsychological tests scores for the study cohort by classification

	Control (n=360)		aMCI (n=52)		naMCI (n=26)		>MCI (n=14)		Low Mood (n = 20)		p value
	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	
Age (years)	75.0 (7)	216.68	78 (10)***	299.92	79.0 (6)*	299.19	80.0 (10)**	350.93	76.0 (9)	266.70	<0.001
Gender (% female)	45.0	-	50.0	-	61.5	-	64.3	-	55	-	0.28
Ethnicity (% White)	99.7	-	98.1	-	92.3	-	92.9	-	100	-	-
Education (years)	12.0 (4)	251.95	11.0 (2)**	187.11	11.5 (3)	229.00	11.0 (1)	155.61	10.0 (2)**	153.20	<0.001
NART IQ[†]	116.0 (13)	252.36	112.0 (19)	203.16	111.5 (22)	186.31	106.0 (13)	172.54	106.0 (15)**	136.72	<0.001
GDS[¶]	1 (2)	196.70	1 (1)[$]	232.16	2 (2)**[$]	293.74	2 (3)[$]	285.19	7 (3)***	428.50	<0.001
CVLT SD Free Recall, z score	0.5 (2)	279.33	-2.0 (1)***	37.81	-0.5 (1)**	180.00	-1.5 (1)***	51.93	-0.75 (2)*[#]	184.75	<0.001
CVLT LD Free Recall, z score	0.0 (2)	279.16	-2.0 (1)***[$]	42.39	-1.0 (1)***	157.85	-2.0 (1)***	71.57	-0.75 (3)*	190.98	<0.001
Brixton Errors[‡^]	19 (9)	210.42	23 (12)**	276.95	31 (12)***	345.00	22.5 (17)	279.50	23.0 (11)*	298.16	<0.001
Trails A[¥], percentile	50 (50)	259.92	30 (40)***	169.44	20 (30)***	127.21	20 (20)**	117.39	40 (58)	193.58	<0.001
Trails B[$], percentile	60 (40)	257.43	25 (40)***	146.70	10 (10)***	80.60	10 (30)**	99.83	30 (50)	173.50	<0.001
VOSP-Incomplete Letters score[†]	20 (1)	245.35	19.5 (1)	229.39	19 (2)	190.60	19 (2)	178.14	19 (1)	184.82	0.017
VOSP-Object Decision score[^]	18 (2)	253.35	18 (4)*	181.86	17.5 (3)	191.63	16.5 (3)*	141.43	17 (3)	200.08	<0.001
VOSP-Dot Counting score	10 (0)	250.19	10 (1)***	188.79	10 (1)***	168.31	10 (1)	203.14	10 (0)	226.08	<0.001

	Control (n=360)		aMCI (n=52)		naMCI (n=26)		>MCI (n=14)		Low Mood (n = 20)		p value
	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	
VOSP-Number Location score[^]	10 (1)	253.86	9 (2)	219.53	8 (3)***	136.27	8 (4)	158.82	8 (6)**	152.80	<0.001
Clock Drawing Test score[†]	5 (0)	247.68	5 (0)[%]	219.60	5 (1.3)**	177.69	4 (2)***	117.18	5 (0)[%]	227.90	<0.001
Graded Naming Test score[€^]	24 (4)	263.61	20.0 (9)***	143.31	19.5 (7)***	149.48	18.5 (12)***	104.43	21.0 (8)	179.88	<0.001
Pyramids & Palm Trees Test score[†^]	51 (2)	257.65	50.5 (2)***	179.28	50.5 (3)**	172.81	50 (2)**	132.29	50.0 (2)**	149.52	<0.001
BADLS score[†]	0 (0)	214.23	0 (3)**[%]	278.62	0.5 (3)*[%]	285.25	13 (4)***	463.18	1.5 (2)*[%]	292.85	<0.001

KEY: BADLS = Bristol Activities of Daily Living Scale; CVLT LD = California Verbal Learning Test, long delay recall; CVLT SD = California Verbal Learning Test, short delay recall; IQR = interquartile range; VOSP = Visual Object and Space Perception

*$p < 0.05$, **$p \leq 0.01$, ***$p < 0.001$ for difference from controls; [#]$p < 0.05$ for difference from aMCI and >MCI; [%]$p < 0.05$ for difference from >MCI; [$]$p < 0.05$ for difference from Low Mood with Bonferroni correction for multiple comparisons

[†]n = 359 control; [¶]n = 330 control, 50 aMCI, 25 naMCI, 13 >MCI; [‡]n = 355 control, 49 aMCI, 24 naMCI, 12 >MCI, 19 low mood; [¥]n = 358 control; [§]n = 356 control, 50 aMCI, 24 naMCI, 9 >MCI, 19 low mood; [€]n = 51 aMCI

NB: Distributions of characteristics/scores were not similar for all groups, therefore comparisons are based on the mean rank (unless indicated [^] where distributions were similar, therefore comparisons are based on the median)

Discussion

The purpose of this chapter was to report our experience of recruiting older people from the community to take part in a study investigating the validity of brief cognitive tests for identifying people with MCI, as well as provide some indication of the rates of MCI within a UK community-based population.

Although the current study was not designed from the outset as a prevalence study, it does provide some indication of the prevalence of MCI within a defined population and thus it is interesting to discuss the findings in relation to previous studies of MCI prevalence. As discussed in Chapter 1 of the thesis, previous studies have shown enormous variation in MCI prevalence, ranging from 3 to 42%, with the variability largely being attributed to a lack of consensus in criteria choice and implementation (Ward et al., 2012). Many of the studies included in the review by Ward and colleagues (2012) were large population-based studies, with the median sample size across all studies exceeding 1000 subjects. Such large studies tend to use abbreviated testing schedules or retrospectively apply criteria to previously assessed patients, resulting in less precise classifications.

The current study included a well characterised, prospectively assessed cohort of older people recruited from the community and a prevalence of 16.5% of MCI was found within the assessed cohort. This finding lies within the range of that reported by other European-based studies applying similar criteria. For instance, a study by Artero and colleagues (2006), conducted in France using a similar GP practice based recruitment strategy, found an almost identical prevalence rate of 16.6% of MCI in their sample (Artero, Petersen, Touchon, & Ritchie, 2006). They also applied the criteria and diagnostic flowchart proposed by Petersen et al (Petersen, 2004; Winblad et al., 2004)

and assessed similar cognitive domains including episodic memory, attention, visuospatial ability and language, albeit using a different, computerised neuropsychological battery of tests. Our finding also falls within the range estimated by Luck and colleagues of 14.1% – 16.6%, which they report in their German study of prevalence of MCI among primary care patients (Luck et al., 2007). Again, although this study used a different battery of neuropsychological tests to those used in the current study, similar cognitive domains were assessed and the same criteria were applied.

In terms of subtypes of MCI, we found that aMCI was twice as prevalent at naMCI. Similar findings have been reported in previous studies, for example the large population-based Mayo Clinic Study of Aging reported a 2:1 ratio of aMCI to naMCI in their sample of almost 3000 participants (Petersen et al., 2010).

It is important to note that, for their prevalence estimates, Artero et al (2006) and Luck et al (2007) only included people with subjective cognitive complaints in their MCI groups. For comparison, however, Luck et al reported additional estimates of MCI prevalence including people without memory complaints. In this case, their estimate of prevalence increased to 23.7 – 26.7%, which is somewhat higher than our estimate (which included those with and without memory complaints). This discrepancy could be explained by the fact that their cognitive domain impairment definition was set slightly lower than ours at 1SD below norm.

The adherence to the criterion of subjective cognitive complaint for the classification of MCI is a matter under debate (Mitchell, 2008b). Although subjective memory complaints are believed to indicate a decline in objective memory performance

(Petersen, 2004), studies have demonstrated poor correlations between the two (Jungwirth et al., 2004; Lenehan, Klekociuk, & Summers, 2012). In the study by Luck and colleagues (2007), it was reported that almost 40% of the cognitively impaired participants did not fulfil the criterion of a subjective cognitive complaint. In the current study, 13% of people classified as having MCI did not report a subjective memory complaint. Conversely, amongst those people who did report a subjective memory complaint, the rate of people who did not display impairment on objective testing was high, indicating that there was a high proportion of the "worried well", a term which has been used previously in the literature to describe people with subjective memory complaints but no objective memory impairment (Ahmed, Mitchell, Arnold, Dawson, et al., 2008), within our sample. The low specificity of memory complaint status for aMCI of 23% reflects this finding. Together these findings provide further evidence to suggest that subjective memory complaints may not be a very reliable indicator of objective cognitive status, and therefore perhaps should not be an essential criterion in the classification of MCI.

Although the study provides an estimate of prevalence of MCI, it should be pointed out that these figures are based on only 6.2% (472/7618) of the contacted population. Without information on the characteristics of the "unassessed" population it is impossible to evaluate whether or not they were similar to the assessed population and would have resulted in similar prevalence estimates. Non-responders or participants who declined to take part may well have been more cognitively impaired than those people who responded and agreed to take part in the study. In fact, a study that compared the prevalence of MCI between people who responded at first contact (quick responders) to a community based survey and those who responded after a follow-up (delayed responders), found a 2.3 fold increase in MCI in delayed compared to quick

responders (Miyamoto et al., 2009). This suggests that any future studies employing similar postal based recruitment strategies should consider re-contacting initial non-responders as the rate of MCI may in fact be higher within that delayed responder cohort.

An important aim of this work was to provide future researchers with information on recruitment rates of MCI. Over a two and a half year recruitment period, during which 7618 people were contacted, 78 people with MCI were recruited. Or, in other words, 1.02% of the contacted population were classified as having MCI and it is this more conservative MCI recruitment rate estimate that should be considered when designing future studies applying similar recruitment techniques. The target at the outset of the study was to recruit 200 people with aMCI; only a quarter of this target was achieved (n=52). This was largely due to the fact that the prevalence of aMCI within people reporting subjective memory impairment was lower than expected (12.5% compared with a predicted 30%, based on studies by Benito-Leon et al (2010) and Mitchell et al (2008)). In addition, the response rate to the flyers was much lower than anticipated at approximately 19%. As discussed previously, it perhaps would have been useful in the current study to have followed up those people who didn't respond initially as this may have resulted in an increased recruitment of individuals with MCI.

Of those people that did respond, only 58.0% were eligible for further follow-up. During the initial phase of the study, one of the most common reasons for exclusion of responders was having a positive depression screening answer to the two Whooley questions (Whooley et al., 1997). Removing this eligibility criterion increased the eligibility rate and further analysis of the Whooley questions revealed that although they were 100% sensitive at picking up depression (as defined by the GDS) they performed

with relatively low specificity, with a false positive rate of 28%. This finding is in line with that from a recent meta-analysis which found a pooled sensitivity of 95% (95% CI 88 to 97) and pooled specificity of 65% (95% CI 56 to 74) for the screening questions (Bosanquet et al., 2015).

Following the removal of the depression screening questions, the main reason for exclusion of responders was having no informant available. This was a requirement for the study due to the use of an informant-rated ADL scale. There is no consensus regarding the assessment of ADLs to satisfy the criterion of preserved ADLs for the classification of MCI. Several options exist including performance-based tasks and questionnaires or interviews (Gold, 2012). Both self-report and informant-report questionnaires are available, however, the practice of using an informant's perspective when rating ADLs is important because there is evidence for impaired insight in MCI (Vogel et al., 2004). Performance based measures could prevent the need for an informant and thereby increase the inclusion rate for future studies. However, they present disadvantages in the fact that they represent a single evaluation point with typical everyday cues removed and also can be time consuming and expensive to administer (Gold, 2012). Questionnaires appear to be the most convenient option for large scale community based MCI case finding, due to their ease of use. However, due to the impact that the requirement of having an informant available has on recruitment rates, self-report ADL questionnaires should be considered for future study designs. It would be particularly useful to conduct a head-to-head comparison of self- and informant-reported ADL questionnaires in MCI cohorts in the future so that the reliability of each method may be evaluated.

Conclusion

The current study has provided a prevalence estimate for MCI of 16.5% within a UK-community based sample of people aged 70 years and above recruited using flyers posted from their GP practice. This estimate is similar to other European-based studies which have used similar recruitment techniques and criteria. However, when taking into account all people initially approached, this rate drops to 1.02% and this more conservative estimate should be considered when designing future studies applying similar recruitment techniques. The main reason for exclusion of potential participants was having no informant available and this fact should be considered in future studies which use an informant-based assessment of ADL ability.

Chapter 4: Assessing the Validity of Two Brief Cognitive Tests for Detecting aMCI in a Community Cohort

Introduction

Amnestic MCI (aMCI), which is associated with elevated rates of conversion to dementia caused by Alzheimer's disease (AD) (Petersen et al., 2001) is largely unrecognised in primary care since its diagnosis depends on complex neuropsychological assessment methods not usually available in this setting. There is a need for simple, quick and sensitive cognitive tests that will provide a more efficient way of identifying people with aMCI. These would provide a useful resource to busy primary healthcare staff who are encouraged, as stated in UK national guidance, to refer people who show signs of MCI for further assessment by memory assessment services to aid early identification of dementia (NICE, 2006). They could also be applied by researchers to find suitable participants for enrolment into studies of candidate interventions targeted at this early stage of cognitive decline.

As reported in the systematic review described in Chapter 2 (recently accepted for publication, (Ozer, Young, Champ, & Burke, 2016)) over 40 brief cognitive tests have been developed and tested to identify people with aMCI. Several of these cognitive tests demonstrated promising diagnostic test accuracy results, though the majority of studies were found to be at a high risk of bias due to the method of participant selection employed. Most studies selected patients with known aMCI from memory clinics and compared their performance on the test under evaluation with an opportunistically recruited group of people assumed to have no cognitive impairment. This exposed the studies to risk of unblinding of the patient assessment process and potentially exaggerated diagnostic accuracy (Lijmer et al., 1999; Whiting et al., 2004). The current

study aimed to address this limitation by assessing the validity of two brief cognitive tests in a cohort of participants all recruited from the community, without prior knowledge of their cognitive status, thereby reducing the risk of bias in the assessment process.

The Memory Alteration Test (M@T) and the Test Your Memory (TYM) test were selected for investigation in this study. The findings from the systematic review in Chapter 2 (conducted in 2013) suggested that other tests, such as the MoCA and AVLT, demonstrated better diagnostic test accuracy than the chosen tests for this study (particularly than the TYM). However, at the time of test selection, which took place in 2011 as part of the NIHR grant application process, the M@T and TYM were deemed to be the most promising tests available for community-based identification of aMCI. The M@T, which is a brief, interviewer-administered memory task, had been highlighted in a review by Lonie and colleagues (2009) as being particularly well suited to use in general practice, with a relatively short administration time (<7 minutes) and simple scoring method and its developers reported it to have very high sensitivity (96%) and high specificity (70-79%) for discriminating between people with aMCI and healthy controls (Rami et al., 2010; Rami et al., 2007). The TYM was identified as another potentially suitable, simple cognitive test for identifying aMCI in primary care. The developers of TYM reported that it had very high sensitivity (93%) and high specificity (86%) for discriminating between people with and without mild Alzheimer's disease (Brown et al., 2009). A subsequent study using a Japanese version of the test highlighted its potential for use as a screening tool for aMCI, reporting high sensitivity (76%) and specificity (74%) (Hanyu et al., 2011). The TYM has the added advantage of being self-administered and requiring minimal supervision.

The aim of the current study was to evaluate the effectiveness of the M@T and TYM for identifying people with aMCI by investigating: (1) their sensitivity/specificity in detecting aMCI in a community-based population in comparison with the widely used standard for diagnosing aMCI based on the Petersen criteria (Petersen, 2004); (2) their test-retest reliability performance; and, (3) their clinical utility, assessed in terms of administration time and completion rates.

Methods

Recruitment

Recruitment methods for the study are covered in detail in Chapter 3. In brief, older people (aged 70 years and above) were contacted via a study information flyer sent to them from their GP practice. Eligible respondents were then contacted by telephone to be invited to take part in the study. Those people who wanted to take part were subsequently enrolled onto the study and gave informed written consent to participate during a visit to their home.

Assessment Procedures

All participants were assessed using the two brief cognitive tests under investigation (M@T and TYM), as well as the standardised battery of neuropsychological tests used to classify their cognitive status (described in Chapter 3). Following the classification flowchart (see Figure 3.1), participants were classified as having (1) amnestic MCI (aMCI), (2) non-amnestic MCI (naMCI), (3) cognitive difficulties beyond MCI (>MCI) and (5) low mood. Those participants who did not meet the criteria for aMCI, or the other possible study classifications, were classified as “controls” and formed the reference group for the subsequent discriminatory analyses.

The M@T and TYM were administered in a randomised order to avoid "order effects", which refers to the differences in participant responses that can result from the order in which tests are presented to them ("Psychology Research and Reference," 2016). The allocation of the first administered test was determined by blocked randomisation (1:1 ratio) in order to limit any potential differences in sample size between groups. Block sizes were varied randomly between 10 and 20. A randomisation website was used to generate the block sizes as well as the allocation sequence for each block (http://www.random.org/lists/).

The first allocated brief cognitive test was administered during Session 1 which took place in the participant's home. The other brief cognitive test followed by the neuropsychological test battery (described in Chapter 3), were administered during Session 2 which took place in the Clinical Research Facility based at Bradford Royal Infirmary within two weeks of Session 1. The M@T, TYM and neuropsychological battery were all administered by research assistants who were blinded to each other's assessments. Both the M@T and TYM were timed by stopwatch. Classifications of participants were agreed in consensus with the study neuropsychologist (Dr Krist Noonan), who was blinded to the results of the M@T and the TYM.

To assess test-retest reliability, Session 3 was arranged for a sample of participants who were re-administered the brief cognitive test they had completed during Session 1. The first 25 consecutive aMCI and control participants that were allocated the M@T first and that were allocated the TYM first were invited to take part in Session 3. When participants refused or could not be contacted for follow-up, the next eligible participant was invited. In cases where participants were part of a couple, both participants were invited to take part in Session 3 so that the reason for their invitation did not have to be

revealed, since cognitive status classifications were not disclosed to participants. Session 3 was scheduled to take place at home within four weeks of Session 1.

Memory Alteration Test (M@T)

The Memory Alteration Test (M@T) (Rami et al., 2007) is an interviewer-administered test comprising a minimum of 33, and a maximum of 43, questions depending on free recall success (see Appendix 4.1). It assesses five cognitive skills (encoding, orientation, semantic memory, free recall and cued recall, with recall intervals of <10 minutes) with a maximum total score of 50. It was developed and validated in Spain but has been translated into English, although not validated in this form. The translated version from the development paper was applied (Rami et al., 2007) (with slight amendments made to the wording of some of the semantic memory questions; see Appendix 4.2).

Test Your Memory Test (TYM)

The Test Your Memory (TYM) test (Brown et al., 2009) is a supervised, self-completed questionnaire comprising ten cognitive tasks, providing assessment of a wider range of cognitive domains than is covered in the M@T (see Appendix 4.3). In addition to memory and orientation tasks, the TYM also includes calculation, fluency, similarities, naming, and visuospatial tasks. As with the M@T, the recall interval for the memory task is <10 minutes. A score out of five is also given for the amount of help that the participant required to complete the task, with higher scores indicating that less support was required. The maximum total score is 50.

Statistical Analysis

All analyses were performed using SPSS Statistics v22 (IBM). Between-group differences (aMCI vs. control) in age, years of education, IQ and the M@T and TYM scores were explored using the Mann-Whitney U test (since the data were non-normally distributed). Difference in gender proportion between the groups was analysed using the X^2 test.

Receiver operating characteristic (ROC) curve analysis was applied to assess the ability of the M@T and TYM global and subtest scores to discriminate between the aMCI group and the control group for a range of cut-off values. The area under the curve (AUC) was reported as a single measure of overall accuracy. Optimal cut-off points were defined as those providing the highest Youden index, which is a way of combining sensitivity and specificity into a single measure, calculated as "sensitivity + (specificity-1)" (Youden, 1950). Positive and negative predictive values and likelihood ratios were calculated for each optimal cut-off point. Wilson's 'score' method (Wilson, 1927) was used to estimate 95% confidence intervals and they were calculated using an online calculator provided by PEDro, the Physiotherapy Evidence Database (Herbert, 2013).

Test-retest reliability of the M@T and TYM was investigated using the established techniques of Bland and Altman (Bland & Altman, 1986). The mean difference between original test and retest scores was calculated, as was the reliability coefficient, which is twice the standard deviation of the differences and provides a measure of random error. Paired sample t-tests were applied to explore for any significant differences between original test and retest scores (since the data were normally distributed). Agreement between original test and retest classifications (based on optimal cut-off scores) was also explored using the kappa statistic to measure agreement beyond that which would

be expected by chance alone (established categories for interpreting the kappa statistic were applied from poor (<0.00) to moderate (0.41 -0.60) to almost perfect (0.81 – 1.00) (Landis & Koch, 1977).

Results

The characteristics of the recruited participants are described in detail in Chapter 3. Of the 472 participants who were recruited and classified, 52 people had aMCI and 360 people were designated controls and formed the reference group for subsequent analyses. Seventy two per cent of these participants completed Session 2 within two weeks of Session 1.

The aMCI participants were significantly older, had fewer years of education and a lower NART IQ than the controls (see Table 3.6). Since these factors could have had an influence on the M@T and TYM scores, age-, education- and IQ-matched controls were randomly selected for the discriminatory analyses. The aim was to select three matched controls for each aMCI participant; however some aMCI participants had less than 3 matches (and some none at all), which resulted in 40 aMCI cases matched with 112 controls. The demographic characteristics of these matched participants are provided in Appendix 4.4.

M@T Performance

Validity

Participants with aMCI scored significantly lower on the M@T than the control participants (35 (10) vs. 45 (5), U = 1459, z = -9.7, p<0.001, see Table 4.1). The box plots demonstrate the distribution of the M@T scores for each group (see Figure 4.1a).

Figure 4.2a shows the ROC curve of the M@T for differentiating the aMCI participants from the matched controls. The AUC was 0.91 and a score of 40 provided the optimal cut-off for discriminating between aMCI and controls (sensitivity 85% (95% CI 70 – 93%), specificity 84% (95% CI 76 – 89%); see Table 4.2). To give context to these figures, the findings indicate that if 100 people with aMCI were administered the M@T, 85 of those people would be correctly classified as having aMCI (i.e. true positives) and the remaining 15 people would be incorrectly classified as not having aMCI (i.e. false negatives). On the other hand, if 100 controls were administered the M@T, 84 of those people would be correctly classified as not having aMCI (i.e. true negatives) and the remaining 16 people would be incorrectly classified as having aMCI (i.e. false positives).

At the developer-recommended cut-off of 37 (Rami et al., 2007), a lower sensitivity (64%) but higher specificity (96%) was achieved. In other words, if the same groups of people in the above example were administered the M@T with this revised cut-off, there would be fewer true positives identified (64 compared with 85), but more true negatives identified (96 compared with 84).

The diagnostic utility parameters for the M@T subtests are also summarised in Table 4.2. The most sensitive subtests to discriminate between the aMCI and control groups were Free Recall and Cued Recall, which both demonstrated AUC >0.85. Orientation was the least sensitive subtest, with the lowest AUC value (0.61).

Reliability

Twenty-five aMCI cases and 31 controls were reassessed with the M@T. Three quarters (75%) of these participants were reassessed within 4 weeks of Session 1. Participants tended to score higher in Session 3 than Session 1 (mean difference 2.8 points (95% CI 2.0 to 3.7); see Table 4.3). This difference was significant ($t(54) = -6.05$, $p<0.001$). The kappa value was 0.54 (indicating "moderate" agreement between sessions).

Utility

The median time to complete the M@T in the control group was 6min 5sec (IQR: 1min 40sec). Participants with aMCI took significantly longer than the control group, with a median time of 8min 15sec (± 1min 45sec) ($U = 2798$, $z = -7.95$, $p<0.001$).

The majority (n=409, 99%) of aMCI and control participants completed all the M@T questions. One aMCI participant had one missing item from the Free Recall subset and two control participants each had one missing item from the Cued Recall subset. These participants were excluded from any analyses concerning these scores (i.e. M@T total score, Free Recall subset and Cued Recall subset discriminatory analyses). In addition, one aMCI participant had missing data from their re-test M@T (three missing items from the Cued Recall subset) and was therefore excluded from the test-retest reliability analyses.

The M@T requires the participant to encode and recall five words: cherry, axe, elephant, piano and green. However, it was noted that the words "axe" and "green" were commonly misheard and had to be repeated.

TYM Performance

Validity

Participants with aMCI scored significantly lower on the TYM than the control participants (41 (8) vs. 47 (4), U = 2921.5, z = -8.1, p<0.001, see Table 4.1). The box plots demonstrate the distribution of the TYM scores for each group (see Figure 4.1b).

Figure 4.3b shows the ROC curve of the TYM for differentiating the aMCI participants from the matched controls. The AUC was 0.80 and a score of 43 provided the optimal cut-off for discriminating between aMCI and controls (sensitivity 63% (95% CI 47 – 76%); specificity 87% (95% CI 79 – 93%); see Table 4.2). To give context to these figures, the findings indicate that if 100 people with aMCI were administered the TYM, 63 of those people would be correctly classified as having aMCI (i.e. true positives) and the remaining 27 people would be incorrectly classified as not having aMCI (i.e. false negatives). On the other hand, if 100 controls were administered the TYM, 87 of those people would be correctly classified as not having aMCI (i.e. true negatives) and the remaining 13 people would be incorrectly classified as having aMCI (i.e. false positives).

At the commonly used cut-off of 44 (Hanyu et al., 2011; Munoz-Neira, Henriquez Chaparro, Delgado, Brown, & Slachevsky, 2014; Szczesniak, Wojtynska, & Rymaszewska, 2013), a slightly higher sensitivity (65%) but lower specificity (80%) was achieved. In other words, if the same groups of people in the above example were administered the TYM with this revised cut-off, there would be more true positives identified (65 compared with 64), but fewer true negatives identified (80 compared with 87).

The diagnostic utility parameters for the TYM subtests are summarised in Table 4.2. All subtests (except for the Fluency subtest) performed with less sensitivity than the global TYM test, and all subtests demonstrated AUC values of less than 0.75. Fluency and Free Recall were the most accurate subtests, with AUCs of 0.72.

Reliability

Nineteen aMCI cases and 30 controls were reassessed with the TYM. The majority (88%) of these participants were reassessed within four weeks of Session 1. Participants tended to score higher in Session 3 than Session 1 (mean difference 1.9 points (95% CI 1.0 to 2.8); see Table 4.3). This difference was significant ($t(48) = -4.40$, $p<0.0005$). The kappa value was 0.51 (indicating "moderate" agreement between sessions).

Utility

The median time to complete the TYM in the control group was 7min 19sec (± 2min 30sec). Participants with aMCI took significantly longer ($p<0.005$) than the control group, with a median time of 9min 26sec (± 2min 32sec) ($U = 4374.5$, $z = -5.59$, $p<0.001$).

Fully completed TYM questionnaires were obtained for all of the aMCI participants and almost all of the controls. Only one control participant had missing items with four missing items from the Orientation subtest and no score for the Help Given subtest. This participant was excluded from analyses concerning these scores (i.e. TYM total score, Orientation subset and Help Given subset discriminatory analyses).

Table 4.1: M@T & TYM test scores for aMCI and Control participants

	aMCI (n=52)	Control (n=360)	p value
M@T global score	35 (10)[‡]	45 (5)[§]	<0.001
TYM global score	41 (8)	47 (4)[†]	<0.001

Data are presented as median (interquartile range) unless indicated otherwise; p value for Mann-Whitney-U test

[‡]n = 51; [§]n = 358; [†]n = 359;

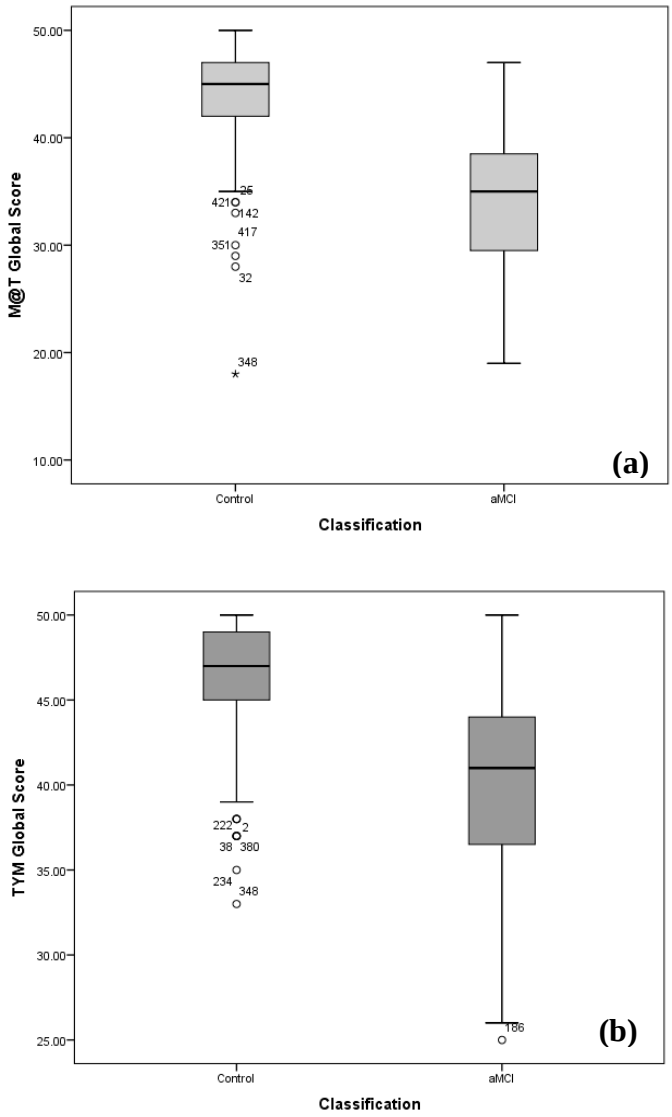


Figure 4.1: (a) M@T and (b) TYM global scores for control and aMCI participants

KEY: o = mild outliers (i.e. data points that lie >1.5 * IQR away from the lower/upper quartile); * = extreme outliers (i.e. data points that lie >3 * IQR away from the lower/upper quartile); IQR = interquartile range

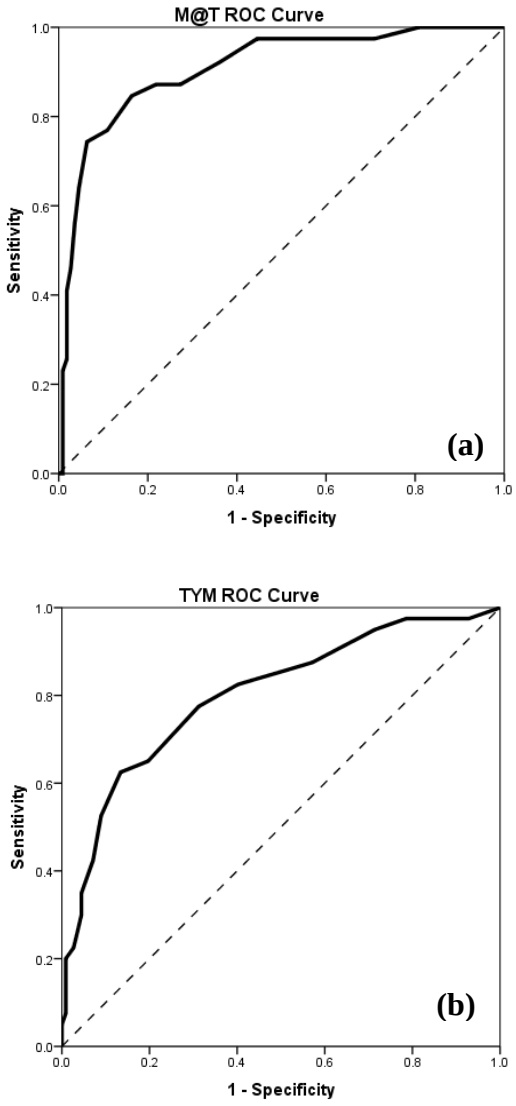


Figure 4.2: Receiver operating characteristics of the (a) M@T and (b) TYM for differentiating aMCI participants from age, education and IQ-matched controls

Table 4.2: Diagnostic utility of M@T to discriminate between aMCI and age, education and IQ-matched controls

Test (maximum scores)	AUC (95%CI)	p value	Optimal Cut-off*	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	LR+ (95%CI)	LR- (95%CI)
M@T Total Score (50)	**0.91 (0.85 – 0.96)**	**<0.001**	**<40**	**85 (70 – 93)**	**84 (76 – 89)**	**34 (25 – 44)**	**98 (96 – 99)**	**5.17 (3.32 – 8.05)**	**0.18 (0.09 – 0.39)**
Encoding (10)	0.79 (0.70 – 0.87)	<0.001	<9	58 (42 – 71)	88 (80 – 92)	32 (22 – 43)	95 (93 – 97)	4.60 (2.63 – 8.04)	0.48 (0.34 – 0.70)
Orientation (5)	0.61 (0.50 – 0.71)	0.05	<5	38 (24 – 53)	83 (75 – 89)	18 (11 – 28)	92 (90 – 95)	2.21 (1.25 – 3.92)	0.75 (0.58 – 0.97)
Semantic (15)	0.74 (0.65 – 0.83)	<0.001	<14	85 (71 – 93)	54 (44 – 63)	15 (11 – 21)	97 (94 – 99)	1.83 (1.44 – 2.32)	0.28 (0.13 – 0.60)
Free Recall (10)	0.88 (0.82 – 0.94)	<0.001	<6	90 (76 – 96)	77 (68 – 84)	28 (21 – 36)	99 (97 – 99)	3.87 (2.72 – 5.50)	0.13 (0.05 – 0.34)
Cued Recall (10)	0.86 (0.78 – 0.93)	<0.001	<9	80 (65 – 90)	78 (70 – 85)	27 (20 – 35)	97 (95 – 99)	3.67 (2.49 – 5.40)	0.26 (0.14 – 0.48)

*cut-off providing highest Youden index; PPV = positive predictive value; NPV = negative predictive values; PPV and NPV calculated for 10% prevalence of MCI (calculated as PPV = TP/(TP+FP); NPV = TN/(TN+FN), where TP = true positive, FP = false positive, TN = true negative, FN = false negative); LR+ = positive likelihood ratio; LR- = negative likelihood ratio

Table 4.3: Diagnostic utility of TYM to discriminate between aMCI and age, education and IQ-matched controls

Test (maximum scores)	AUC (95%CI)	p value	Optimal Cut-off*	Sensitivity	Specificity	PPV	NPV	LR+	LR-
TYM Total Score (50)	**0.80 (0.72 – 0.88)**	**<0.001**	**<43**	**63 (47 – 76)**	**87 (79 – 93)**	**32 (22 – 43)**	**95 (93 – 97)**	**4.67 (2.75 – 7.92)**	**0.43 (0.29 – 0.65)**
Orientation (10)	0.57 (0.46 – 0.67)	0.22	<10	30 (18 – 45)	83 (75 – 89)	15 (9 – 24)	91 (89 – 94)	1.77 (0.95 – 3.31)	0.84 (0.68 – 10.5)
Copying (2)	0.52 (0.41 – 0.62)	0.77	<1	5 (1 – 16)	99 (95 – 100)	33 (1 – 70)	91 (88 – 94)	5.60 (0.52 – 60.1)	0.96 (0.89 – 1.03)
Semantic (3)	0.67 (0.56 – 0.77)	0.002	<2	40 (26 – 55)	91 (84 – 95)	31 (20 – 44)	93 (91 – 96)	4.48 (2.22 – 9.05)	0.66 (0.51 – 0.85)
Calculation (4)	0.58 (0.47 – 0.69)	0.14	<4	40 (26 – 55)	73 (64 – 80)	13 (8 – 20)	92 (89 – 95)	1.49 (0.92 – 2.43)	0.82 (0.62 – 1.08)
Fluency (4)	0.72 (0.63 – 0.82)	<0.001	<4	73 (57 – 84)	66 (57 – 74)	18 (13 – 24)	96 (93 – 98)	2.14 (1.55 – 2.95)	0.41 (0.25 – 0.70)
Similarities (4)	0.61 (0.51 – 0.72)	0.04	<4	53 (38 – 67)	68 (59 – 76)	14 (9 – 20)	93 (90 – 96)	1.63 (1.10 – 2.43)	0.69 (0.49 – 0.99)
Naming (5)	0.54 (0.43 – 0.65)	0.46	<5	13 (5 – 26)	96 (90 – 98)	24 (11 – 45)	91 (88 – 94)	2.8 (0.86 – 9.2)	0.91 (0.80 – 1.04)
Visuospatial 1 (3)	0.50 (0.40 – 0.61)	0.99	<1	13 (5 – 26)	94 (88 – 97)	17 (8 – 34)	91 (88 – 94)	2.00 (0.67 – 5.95)	0.93 (0.82 – 1.06)

Visuospatial 2 (4)	0.53 (0.42 – 0.63)	0.65	<4	15 (7 – 29)	90 (83 – 94)	13 (6 – 26)	91 (88 – 94)	1.53 (0.60 – 3.86)	0.94 (0.82 – 1.09)
Free Recall (6)	0.72 (0.62 – 0.82)	<0.001	<3	50 (35 – 65)	93 (87 – 96)	42 (29 – 56)	94 (92 – 97)	7.00 (3.35 – 14.6)	0.54 (0.39 – 0.74)
Help (5)	0.53 (0.43 – 0.64)	0.55	<4	13 (5 – 26)	93 (86 – 96)	15 (7 – 31)	91 (88 – 94)	1.75 (0.61 – 5.04)	0.94 (0.83 – 1.07)

*cut-off providing highest Youden index; PPV = positive predictive value; NPV = negative predictive values; PPV and NPV calculated for 10% prevalence of MCI (calculated as PPV = TP/(TP+FP); NPV = TN/(TN+FN), where TP = true positive, FP = false positive, TN = true negative, FN = false negative); LR+ = positive likelihood ratio; LR- = negative likelihood ratio

Table 4.4: Test-retest reliability of the M@T and TYM

	Bland and Altman Results				**Cohen's Kappa**
	Mean difference	**95% CI for mean difference**	**SD_{diff}**	**Reliability coefficient**	
M@T	-2.8	-2.0 to -3.7	3.5	6.9 (out of 50)	0.54*
TYM	-1.9	-1.0 to -2.8	3.0	6.0 (out of 50)	0.51*

*p<0.001

Discussion

The current study aimed to assess the accuracy of two brief cognitive tests (M@T and TYM) for identifying people with aMCI in the community. The M@T performed with higher diagnostic test accuracy than the TYM, with higher sensitivity (85% vs 63%), similar specificity (84% vs. 87%) and higher overall accuracy as demonstrated by the AUC values (0.91 vs. 0.80). Both tests were associated with a learning effect such that a second assessment repeated within one month of the first showed higher test scores. Both tests were acceptable to participants with completion times of less than ten minutes and very few missing items.

Although the M@T demonstrated reasonably high levels of sensitivity and specificity for aMCI, the study did not reproduce the very high diagnostic test accuracy (DTA) results reported in previous studies. For example, a recent study by Custodio et al reported that a cut-off score of 37 had a sensitivity and specificity of 98% (AUC = 0.999) to differentiate aMCI from controls (Custodio et al., 2014). The developers of the M@T recommend a cut-off score of 37 and they report sensitivity of 96% and specificity of 70-79% at this cut-off (Rami et al., 2010; Rami et al., 2007). However, a higher optimal cut-off value was found for the current sample (<40) and a lower sensitivity (63%) but higher specificity (96%) was demonstrated at the recommended cut-off.

The DTA results for TYM also differed from those reported in previous studies. A cut-off score of 44 has been recommended in three previous studies of TYM (Hanyu et al., 2011; Munoz-Neira et al., 2014; Szczesniak et al., 2013). These studies report sensitivities of 74 – 86% and specificities of 60 – 74% at this cut-off. A slightly lower

optimal cut-off value was found in the current sample (<43) and lower sensitivity (65%) but higher specificity (80%) at the previously recommended cut-off.

These DTA discrepancies might be explained by the community-based recruitment method that was used in the current study. The previous studies were all conducted in secondary/specialist care settings such as memory clinics (Hanyu et al., 2011; Rami et al., 2010; Rami et al., 2007), neurology departments (Custodio et al., 2014; Munoz-Neira et al., 2014) or psychiatry units (Szczesniak et al., 2013) and most recruited their aMCI sample from patients attending clinics and their "control" sample from a separate source, such as other hospital departments or the wider community (Custodio et al., 2014; Hanyu et al., 2011; Rami et al., 2007; Szczesniak et al., 2013). Studies which use a "case-control" design such as this are known to exaggerate diagnostic accuracy (Lijmer et al., 1999; Whiting et al., 2004). The current study used a sampling method that was designed to reflect how the brief cognitive tests might be applied in routine care in the future, that is, community based aMCI case finding. This approach has resulted in more conservative estimates of DTA which are likely to be more generalizable to unselected populations. Assessing all participants with the same reference standard also meant that verification bias was avoided, which occurs when only a proportion of the study population receive confirmation of the diagnosis (usually those with positive test results) and can also result in overestimation of DTA values (Lijmer et al., 1999; Whiting et al., 2004). The fact that both tests were found to perform at lower sensitivity at the recommended cut-offs than previously demonstrated indicates that the current aMCI population were less impaired than those included in previous studies, likely to be a result of the community-based, rather than secondary care-based, approach to recruitment.

As reported by the developers of M@T, the current study also demonstrated that Free Recall and Cued Recall were the most accurate sub-tests for discriminating between the aMCI and control groups. It is perhaps unsurprising that these recall scores are the most useful for identifying aMCI since it is well known that episodic memory is impaired in aMCI and early AD (Petersen et al., 1999). This is thought to be the result of early pathological changes that occur in the hippocampus and medial temporal lobe (Jack, Shiung, Weigand, & al, 2005). Similarly, the Free Recall subtest was found to be the most accurate TYM subtest for identifying people with aMCI. Some TYM subtests were found to be of less value in discriminating between aMCI and controls (e.g. the naming subtest and two visuospatial subtests, all with AUC values ~0.5). The success of the recall subtests at detecting aMCI reflect the findings from the systematic review in Chapter 2, which found that the highest performing tests were those that involved some element of word recall (such as the AVLT-SR, HVLT-LE and FBMS). Together, these findings indicate that the Recall subtest scores are particularly useful for identifying aMCI and that particular emphasis should be placed on these subtests when interpreting scores.

It is important to note here that an algorithmic, rather than clinical, categorisation of aMCI was applied in this study. Although this differs from usual clinical practice, which would involve the incorporation of clinical and neurological examination to make a final diagnosis, it enabled the criteria to be applied in a standardised and objective manner, thereby ensuring reliability of the classifications (Petersen et al., 2014). The M@T and TYM have both demonstrated that they are valid in identifying people with aMCI as classified using this algorithmic method. Of course, in practice, further clinical assessment would be required to make a differential diagnosis and it is the intention that

these instruments would be used as a first screening stage in clinical practice and not as diagnostic tools.

Both the M@T and TYM performed with moderate test-retest reliability. Participants tended to score higher in the second session than the first on both tests indicating that there may have been a learning/practice effect. This is commonly seen with repeated cognitive testing (Heilbronner et al., 2010). The reliability coefficient (which reflects random error) was fairly high for both tests at 6.9 and 6.0 points for M@T and TYM respectively. These values give an indication of the maximum change in score on retest that might be expected by chance in the absence of change in an individual's cognitive status. In other words, only a change of score that is more than 7 points for M@T and 6 points for TYM would represent real change for an individual patient. To the authors' knowledge, this is the first study to provide these data. This is relevant if these tests are to be used in applications such as measuring the effectiveness of interventions or monitoring change in cognition over time (although testing intervals may be longer than four weeks in these instances, which may lessen any practice effects).

Both tests were quick to administer, taking less than 10 minutes, with the M@T being slightly quicker than the TYM (by approximately one minute, on average). Furthermore, there was very little missing data for both tests indicating that there were no issues with administering them. Both tests were designed to be administered by non-specialist staff (with the TYM requiring minimal supervision). A particular issue with the M@T arose concerning the words used to assess episodic memory. The words "axe" and "green" were often misheard by the participants and had to be repeated and so these may need to be replaced by more easily distinguishable words for use in English

speaking populations (e.g. “hammer”, “yellow”). Any adaptations of the M@T would ideally need to be re-validated in further DTA studies.

A limitation of the current study is that there was no long term follow-up of the participants and so the prognostic abilities of the tests cannot be commented on here. Future studies are required to see how accurate the tests are at discriminating between those people who go on to develop dementia and those who remain stable or improve. It would also be interesting to evaluate how the M@T and TYM might perform relative to other commonly used brief cognitive tests, such as the Memory Impairment Screen (MIS) (Buschke et al., 1999) and the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) in a similar setting and head-to-head comparative studies are warranted in the future.

Additionally, it is important to note here that, at the outset of the study, the aim had been to recruit 200 aMCI participants from the community on which to assess the validity of the tests (see Chapter 3). The fact that this sample size was not achieved has limited the precision with which the sensitivity of the tests can be reported here, with the tests demonstrating a lower bound value of sensitivity for aMCI of 70% and 47% for M@T and TYM respectively. Nevertheless, the point estimates of sensitivity provide useful indicators of the expected performance of the tests, but future studies with larger sample sizes would be required to provide more precise estimates.

Conclusion

In summary, the current study has provided evaluation of the performance of M@T and TYM within a community-based UK setting, providing results that are generalizable to the wider population. Amnestic MCI is largely unrecognised in primary care due to the

lack of simple, quick and sensitive cognitive tests. Both M@T and TYM were simple and quick to use and demonstrated moderate test-rest reliability. However, M@T was found to perform with higher DTA than TYM and could provide an efficient and accurate method for identifying aMCI in clinical or research settings.

Chapter 5: Exploring the Use of Reaction Time Measures in Identifying Cognitive Impairment

Introduction

Much research on the early detection of cognitive impairment and dementia has concentrated on the assessment of memory and other cognitive domain performance. However, there is growing interest in the use of other indicators of cognitive function, such as processing speed measures. A meta-analysis conducted to determine the characteristics of cognitive domain impairment in preclinical Alzheimer's disease demonstrated that, in addition to the well-known deficits in episodic memory, losses in processing speed also occur early in the disease process (Backman, Jones, Berger, Laukka, & Small, 2005). A simple way of assessing processing speed is to measure the time it takes to react to a stimulus and this Chapter presents work which was conducted to evaluate the effectiveness of reaction time task derived measures in identifying MCI. These measures were selected as further potentially useful, simple tests that could provide an alternative to the traditionally used memory tests described previously. Measures based on reaction time tasks have an added benefit in that they are unlikely to be influenced by administrator bias, education level and language ability like many of the aforementioned tests.

Previous research has demonstrated that mean reaction times tend to increase in normal ageing (Simon, 1968) and in pathological ageing, with studies showing increased reaction times in patients with cognitive impairment compared to unimpaired controls (Anstey et al., 2007; Dixon et al., 2007). Processing speed is thought to reflect underlying neural integrity and it has been theorised that cognitive performance is degraded when processing speed is slow due to a limited availability of neural resources

to support higher level cognitive behaviour such as episodic memory (Salthouse, 1996). In their study, Anstey et al (2007) provided evidence to suggest that processing speed performance is related to brain structure, with their results demonstrating that faster reactions times were associated with larger corpus callosum size in both healthy controls and those with mild cognitive disorders.

As well as mean level of processing speed, the consistency with which an individual performs across trials within a task has also been suggested as being an important indicator of cognitive functioning (Hultsch & MacDonald, 2004; Jensen, 1992). The term most commonly used to describe a person's level of consistency on a task is intra-individual variability (IIV) and it has been proposed that this measure may provide another behavioural marker of neurobiological disturbance (Hultsch, Strauss, Hunter, & MacDonald, 2008; MacDonald, Nyberg, & Backman, 2006). Previous research has demonstrated that IIV increases in normal ageing (for review see (Dykiert, Der, Starr, & Deary, 2012), traumatic brain injury (Hetherington, Stuss, & Finlayson, 1996; Stuss, Pogue, Buckle, & Bondar, 1994), mild cognitive impairment (Christensen et al., 2005; Dixon et al., 2007; Gorus, De Raedt, Lambert, Lemper, & Mets, 2008) and mild dementia (Burton, Strauss, Hultsch, Moll, & Hunter, 2006; Hultsch et al., 2000). In their study, Hultsch et al (2000) demonstrated that IIV was greater in patients with mild dementia than in cognitively healthy elderly people, regardless of whether or not they had arthritis, indicating that increased IIV is probably primarily due to central neurological rather than somatic disturbances. In addition, Burton et al (2006) demonstrated that IIV is most likely associated with specific, rather than general nervous system disturbances, with their finding that patients with AD were more inconsistent than those with Parkinson's disease. Studies have also shown that IIV in reaction time predicts longitudinal cognitive decline in ageing populations (Bielak,

Hultsch, Strauss, Macdonald, & Hunter, 2010; Lovden, Li, Shing, & Lindenberger, 2007; MacDonald, Hultsch, & Dixon, 2003).

Increases in IIV in reaction time performance have been proposed to reflect the increased neural noise and reduction in cortical representation that could result from the white matter decline that occurs in the brain during normal and pathological aging (MacDonald, Li, & Backman, 2009). Evidence to support this theory has been demonstrated in MRI studies; for instance, Bunce et al. (2007) found that white matter lesioning, particularly in the frontal lobe, was associated with elevated IIV on simple reaction time tasks. In addition, Jackson et al (2012) found strong associations between IIV measures and total cerebral white matter volume, as well as frontal and parietal region volumes, in healthy older adults and participants with early stage Alzheimer's disease. Increased IIV has been linked to frontal cortex mediated processes such as attentional lapses (Bunce, Warr, & Cochrane, 1993) and fluctuations in executive control (West, Murphy, Armilio, Craik, & Stuss, 2002).

Previous evidence suggests a link between cognition and performance on simple reaction time (RT) tasks, leading to the proposal that measures derived from RT tasks have the potential to aid the identification of a range of neurobiological disorders, including mild cognitive impairment (Bunce et al., 2013). This chapter reports the findings from a study that was conducted to investigate the use of RT task derived measures in identifying cognitive impairment.

Studies to date of RT performance in cognitive impairment and dementia have tended to use clinical samples in which the disease is likely to be relatively far progressed (Burton et al., 2006; Hultsch et al., 2000). Alternatively, studies investigating pre-dementia

phases have tended to use a rather broad definition of mild cognitive impairment that includes individuals with “age-associated memory impairment”, “aging-associated cognitive decline” and “mild neurocognitive disorder” (Anstey et al., 2007; Christensen et al., 2005) or have focussed on only the amnestic form of MCI (Gorus et al., 2008). This study aimed to evaluate the use of RT task derived measures within a well-characterised sample of community-dwelling participants including healthy controls, people with amnestic and non-amnestic MCI and people with cognitive difficulties beyond MCI (i.e. possible early dementia). People with low mood were also included since it has been proposed that depression may affect cognitive performance and has been associated with increased IIV in RT performance (Bunce, Handley, & Gaines, 2008).

Participants were administered two RT tasks: (i) a simple, two choice RT task (2CRT), with button box response and (ii) a more complex five choice RT task (5CRT), with touchscreen response. The inclusion of RT tasks with varying complexity enables the influence of task complexity to be investigated. There is evidence to suggest that increasing task complexity is associated with poorer performance in cognitively impaired groups (Dixon et al., 2007; Gorus et al., 2008; Hultsch et al., 2000); however, questions still remain as to the optimal level of complexity of RT task that should be used to assess RT performance in the identification of cognitive impairment (Bielak et al., 2010). Also, the use of touchscreen technology enabled spatial accuracy, as well as speed, of response to be measured, and thus provided an additional novel measure, currently unexplored (to the author’s knowledge) within this field of research. In summary, the current study aimed to investigate the effectiveness of a range of measures derived from RT tasks for identifying people with cognitive impairment, by addressing the following questions: (1) Are there differences in (a) mean RT and RT

variability and (b) accuracy measures between people depending on their cognitive classification?; (2) Does the complexity of the RT task (2CRT vs. 5CRT) have a differential effect on RT performance that is dependent on cognitive classification?; and (3) Are RT task derived measures able to predict cognitive classification?

Methods

Participants

Older people (aged 70 years and above) were invited to participate via flyers sent to them from their GP practice. They were part of a larger study cohort recruited to assess the validity of two brief cognitive tests (described previously in Chapter 4). All participants were evaluated using the battery of cognitive assessments described in Chapter 3 and were categorised into the following groups: (1) Control; (2) Amnestic MCI (aMCI); (3) Non-amnestic MCI (naMCI); (4) Cognitive difficulties beyond MCI (>MCI) and (5) Low mood. A total of 225 people were included in this study. However, one person could not be categorised due to a hearing impairment that impacted on their performance in the cognitive assessments, and was therefore excluded, leaving a total of 224 people in the study.

Participants with cognitive impairment

Thirty-nine people met the Petersen criteria for MCI (Petersen, 2004). Of these, 28 demonstrated impairment in memory (defined as CVLT Short Delay and Long Delay free recall $\geq$ 1.5 standard deviations below mean of published norms (Delis et al., 2000)) and were classified as having aMCI (amnestic MCI). The remaining 11 people demonstrated impairment in other non-memory domain(s) and were classified as having naMCI (non-amnestic MCI). Both single domain and multi-domain MCI participants were included.

Nine people demonstrated cognitive impairment that had detrimental effects on activities of daily living (as measured by the informant-administered Bristol Activities of Daily Living Scale (Bucks et al., 1996)) and were therefore classified as having cognitive impairment beyond MCI (>MCI).

Participants with low mood

Twelve people scored ≥6 on the Geriatric Depression Scale (Sheikh & Yesavage, 1986) and were classified as having a low mood.

Control participants

The remaining 164 people did not meet any of the previously described classifications and were therefore classified as controls and formed the reference group for the subsequent analyses.

The study was approved by the Yorkshire and The Humber National Research Ethics Service Committee (ref: 12/YH/0207) and all participants gave informed written consent. Ethical approval was also granted from the University of Leeds for the administration of the RT tasks, which were developed and administered using University-owned equipment (ref: 13.0256).

Reaction Time Tests

Participants performed two reaction time (RT) tasks. One was a simple two choice RT task and the other was a more complex five choice RT task. Both tasks were designed in ePrime ® version 2.0.8.90 (Psychology Software Tools Inc., USA) and run on a Toshiba Portege M750-116 Touchscreen laptop. Participants were sat positioned an

arm's length (30-50cm) away from the screen when performing the tasks. The order in which the tasks were administered was alternated between participants.

Two Choice Reaction Time Task (2CRT)

This task involved participants responding to a black circular target randomly presented on either the left or right hand side of a white screen (see Figure 5.1a) by pressing the corresponding button on a button box. A CEDRUS ® RB-540 button box attached via USB port to the laptop was used in this task. The circular targets were 25mm in diameter and were presented with an inter-stimulus interval of 500ms, during which a black fixation cross appeared in the centre of the screen. The circular target was positioned 4cm from the fixation cross and stayed on screen until the participant pressed a button. Participants were instructed to respond as quickly and as accurately as possible. Participants completed 12 practice trials, followed by 48 experimental trials. Based on an imposed maximum time limit per trial of 10 seconds, the task did not take longer than 10 minutes to administer, with the majority of participants able to complete the task in under 5 minutes.

Five Choice Reaction Time Task (5CRT)

This task utilised the touchscreen function of the laptop. In this task, participants responded to black circular targets appearing on a white screen by touching the target. The circular targets were 20mm in diameter and were randomly presented around the screen in one of five possible positions, equidistant from the centre at 9cm (see Figure 5.1b). The inter-stimulus interval was 500 or 1000ms (determined randomly) and during this time a box containing the words "touch here" appeared at the bottom of the screen. Participants were instructed to begin each trial by pressing and holding the "touch here" box, ensuring that they began each trial from the same starting point. The circular target

stayed on screen until the participant touched the screen. Again, participants were instructed to respond as quickly and as accurately as possible. Participants completed 5 practice trials, followed by 50 experimental trials. Based on an imposed maximum time limit per trial of 10 seconds, the task did not take longer than 9 minutes and 10 seconds to administer, with the majority of participants able to complete the task in under 5 minutes.

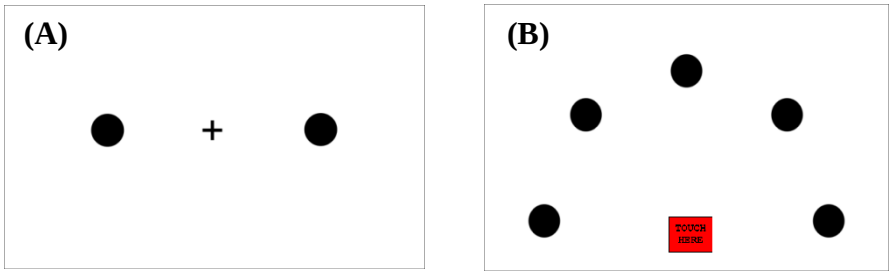


Figure 5.1: Possible target positions on the (A) 2CRT task and (B) 5CRT task

Data Processing

Two Choice Reaction Time Task (2CRT)

The RT in milliseconds (ms) was registered for each trial, and was measured as the time from target onset until a button response was identified. The button that was pressed for each trial was also recorded so that correct and incorrect responses could be identified. The data was filtered so that all RTs <150ms (deemed to be anticipatory rather than genuine responses) and all incorrect responses (i.e. where the participant pressed the wrong button) were excluded. Excessively slow trials (i.e. RTs > individual mean RT + 3SDs) were also removed and replaced with the mean RT over the remaining trials for each participant. These lower and upper bounds have been suggested by previous research (Hultsch, MacDonald, & Dixon, 2002; Hultsch et al., 2000) and removing such outliers and replacing missing values represents a conservative approach to computing IIV (Bunce et al., 2008). As a result of the filtering process, a total of 320 (3.1%) trials were excluded and 173 (1.7%) excessively slow trials were replaced.

Following the filtering of the data, the mean RT and RT IIV were calculated for each participant. Different measures have been proposed in the reporting of IIV (Hultsch et al., 2000). The raw standard deviation (SD) of responses has been used, but it has the disadvantage of being related to mean level of performance. A way of adjusting for this potentially confounding effect is to compute the coefficient of variation (CV), which involves dividing the raw SD by the mean RT (Hultsch & MacDonald, 2004) and it is this measure which has emerged as the standard measure of RT consistency recently due to its relatively easy calculation and high association with other measures of inconsistency (Bunce et al., 2013; Jackson et al., 2012; Lovden et al., 2007). Given this rationale, the CV was therefore calculated for each participant as a measure of IIV in this study. Finally, the number of incorrect responses for each participant was also calculated.

Five Choice Reaction Time Task (5CRT)

The RT in milliseconds (ms) was measured as the time from target onset until a touchscreen response was identified. The RT data was filtered in the same way as for the 2CRT task (see previous section). For this task, incorrect responses/task failures were defined as those trials where the participant touched the screen at a distance of >200 pixels from the target centre (these trials were outside the range of that deemed to be a "genuine" response). Trials were defined as "misses" when the participant touched the screen at a distance of >50 pixels but <200 pixels from the target centre ("misses" were classed as genuine attempts to touch the target and therefore included in the analysis). As a result of the filtering process, a total of 947 (9.1%) trials were excluded and 169 (1.6%) excessively slow trials were replaced.

In addition to mean RT, RT IIV and number of misses, spatial accuracy variables were also calculated for this task. Spatial accuracy was defined as the distance (measured in pixels) of the participant's touch response from the centre of the target. Mean spatial accuracy and spatial accuracy IIV (calculated as the coefficient of variation: spatial accuracy SD/mean spatial accuracy) were calculated for each participant.

Statistical Analyses

All analyses were performed using SPSS Statistics v22 (IBM). Between-group differences in baseline characteristics (including age, years of education, NART IQ, GDS score and the neuropsychological test battery scores) were explored using the Kruskal-Wallis H test (since the data were not normally distributed). Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Difference in gender proportion between the classification groups was analysed using the Chi-square test of homogeneity.

Mean RT and RT IIV for both tasks were also not normally distributed and an inverse transformation was applied to the data to achieve approximately normal distributions prior to subsequent analyses being run. One-way and mixed measures ANOVAs, with classification (cognitive status) as the between-group variable and task complexity (2CRT vs. 5CRT) as the within-group variable were used to explore for differences in mean RT and RT IIV between groups. Bonferroni corrected post hoc group-by-group comparisons were subsequently performed.

All accuracy variables (number of incorrect responses on the 2CRT task, number of misses, mean spatial accuracy and spatial accuracy IIV on the 5CRT task) were also not normally distributed. An inverse transformation was applied to mean spatial accuracy

and spatial accuracy IIV data to achieve approximately normal distributions prior to subsequent analyses being run. One way ANOVAs (with Bonferroni corrected post-hoc pair-wise comparisons) were applied to explore any between group-differences. Due to the nature of the data, the number of incorrect responses on the 2CRT task and the number of misses on the 5CRT could not be transformed and so between group differences were explored using the Kruskal-Wallis H test (with Bonferroni corrected post-hoc pairwise comparisons).

Cohen's d effect sizes for any significant pairwise comparisons were calculated and were interpreted as follows: 0.2 < Cohen's d < 0.5 = small, 0.5 < Cohen's d < 0.8 = medium and 0.8 < Cohen's d = large effect size (Cohen, 1988; Lakens, 2013).

Multinomial regressions were applied to assess whether the RT measures could predict classification. Age and NART IQ were included as covariates and the models were checked to ensure there was no multi-collinearity between the included variables. It is important to note that years of education was not included since it significantly correlated with both age and NART IQ.

Results

Baseline Characteristics

The baseline characteristics and neuropsychological battery test scores of the participants by classification are shown in Table 5.1. Results from the Kruskal-Wallis H tests suggested that there were statistically significant differences between the groups in age ($\chi^2(4) = 13.942$, p=0.007), years of education ($\chi^2(4) = 12.479$, p=0.014) and NART IQ ($\chi^2(4) = 10.755$, p=0.029). However, post-hoc pair-wise comparisons revealed only a

significant difference in age between the controls and >MCI group ($p=0.023$), with the >MCI group being significantly older than the control group.

As the neuropsychological test scores formed the basis of the classifications, as expected, statistically significant differences were found between groups in GDS, CVLT short and long delay recall, Brixton Errors, Trail making tests A and B, VOSP-dot counting and -number location scores, CDT, GNT, BADLS (all at $p<0.001$) and VOSP-object decision score and PPT (at $p<0.05$) (see Table 5.1).

Table 5.1: Baseline characteristics and neuropsychological tests scores for the RT study cohort by classification

	Control (n=164)		aMCI (n=28)		naMCI (n=11)		>MCI (n=9)		Low Mood (n = 12)		p value
	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	
Age (years)	75.0 (8)	104.67	77.5 (11)	125.79	81.0 (6)	144.32	80.0 (13)*	172.22	75.5 (12)	114.54	0.007
Gender (% female)	45	-	43	-	55	-	56	-	33	-	0.823
Education (years)	11 (3)	118.92	10 (2)	84.59	12 (2)	138.09	11 (1)	85.83	10 (2)	86.42	0.014
NART IQ[†]	116.0 (14)	119.74	112.5 (19)	99.27	115.0 (28)	100.27	105.0 (13)	78.56	106.5 (19)	72.42	0.029
GDS	1 (2)	99.94	2 (3)[$]	119.34	2 (2)[$]	131.05	2 (2)	156.06	7.5 (3)***	218.50	<0.001
CVLT SD Free Recall, z score	0.5 (2)	134.98	-2.0 (1)***	19.80	-0.5 (1)	82.18	-1.5 (1)***	24.56	0.25 (3)[#]	115.25	<0.001
CVLT LD Free Recall, z score	0.5 (2)	135.10	-2.0 (1)***	24.91	-1.0 (1)**	54.09	-2.0 (2)***	35.83	0.50 (5)[#]	119.04	<0.001
Brixton Errors[‡^]	18 (8)	97.70	22.5 (11)	131.88	32.5 (17)**	176.75	22.5 (18)	134.56	22.0 (12)	130.23	<0.001
Trails A[†], percentile	60 (50)	127.38	30 (20)***	71.61	10 (20)***	45.41	10 (30)*	57.06	40 (45)	99.62	<0.001
Trails B[§], percentile	60 (50)	123.93	25 (45)***	67.92	10 (15)***	29.90	15 (35)*	47.08	30 (35)*	69.04	<0.001
VOSP-Incomplete Letters score	20 (1)	115.71	20 (1)	117.88	19 (3)	76.27	19 (2)	102.22	19 (1)	97.00	0.204
VOSP-Object Decision score[^]	19 (3)	120.35	18 (3)	99.48	17 (2)	77.36	17 (3)	68.00	18 (5)	101.12	0.018
VOSP-Dot Counting score	10 (0)	121.06	10 (1)**	87.48	10 (2)**	76.82	10 (2)*	79.22	10 (0)	111.58	<0.001
VOSP-Number Location score[^]	9.5 (1)	122.14	9 (3)	102.89	7 (3)**	48.68	8 (4)	88.83	8 (5)	79.42	<0.001

	Control (n=164)		aMCI (n=28)		naMCI (n=11)		>MCI (n=9)		Low Mood (n = 12)		p value
	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	Median (IQR)	Mean Rank	
Clock Drawing Test score[†]	5 (0)	116.72	5 (0)[%]	109.55	5 (2)	95.59	4 (1.5)***	54.89	5 (0)[%]	111.42	<0.001
Graded Naming Test score[^]	24 (4)	126.85	19.5 (8)***	71.95	17 (12)	76.82	19 (12)**	54.00	20 (8)	87.62	<0.001
Pyramids & Palm Trees Test score[^]	51 (2)	122.69	50.5 (2)	88.73	49 (3)	71.45	51 (2)	86.28	50.5 (3)	86.04	0.002
BADLS score[†]	0 (0)	99.68	1 (3)**[%]	137.54	0 (2)[%]	119.23	14 (3)***	219.00	0.5 (2)[%]	132.83	<0.001

KEY: BADLS = Bristol Activities of Daily Living Scale; CVLT LD = California Verbal Learning Test, long delay recall; CVLT SD = California Verbal Learning Test, short delay recall; IQR = interquartile range; VOSP = Visual Object and Space Perception

*$p < 0.05$, **$p \leq 0.01$, ***$p < 0.001$ for difference from controls; [#]$p < 0.05$ for difference from aMCI and >MCI; [%]$p < 0.05$ for difference from >MCI; [$]$p < 0.05$ for difference from Low Mood with Bonferroni correction for multiple comparisons

[†]n = 163 control; [‡]n = 161 control, 26 aMCI, 10 naMCI, 8 >MCI, 11 low mood; [§]n = 160 control, 26 aMCI, 10 naMCI, 6 >MCI

NB: Distributions of characteristics/scores were not similar for all groups, therefore comparisons are based on the mean rank (unless indicated [^] where distributions were similar, therefore comparisons are based on the median)

Exclusions

In the following analyses, for the 2CRT task, data could not be collected from 8 control participants and 2 aMCI participants due to software issues occurring during administration of the task. For the 5CRT task, data could not be collected from 3 control participants, 1 naMCI participant, 1 >MCI participant and 1 low mood participant due to software issues occurring during administration of the task. In addition, the 5CRT task could not be administered with 1 >MCI participant due to comprehension difficulties and had to be abandoned for 1 control participant due to visual difficulties. Finally, 5CRT data was excluded for 9 participants due to them having <50% of trials remaining after the filtering process (6 controls, 1 naMCI, 1 >MCI, 1 low mood). The large failure rate for these participants indicated that they had failed to comply with the task objectives and therefore their data was deemed to be invalid.

Mean Reaction Time

Table 5.2 and Figure 5.2 present the mean reaction times of the participants by classification for each task. Results from the one way ANOVAs suggested that there were statistically significant differences between groups for mean RT on the 2CRT task ($F(4, 209) = 6.731$, $p<0.001$, partial $\eta^2 = 0.114$) and on the 5CRT task ($F(4, 202) = 7.453$, $p<0.001$, partial $\eta^2 = 0.129$). Post hoc analyses revealed that persons with aMCI and naMCI were significantly slower than the controls on both the 2CRT task and the 5CRT task ($p\leq0.01$). Effect sizes for these comparisons were large for both aMCI and naMCI on the 5CRT task (Cohen's d = 1.07 and 1.09 respectively) and medium for aMCI and large for naMCI (Cohen's d = 0.73 and 1.30 respectively) on the 2CRT task.

All groups were faster on the simple task (2CRT) compared to the more complex task (5CRT). This difference was statistically significant for all groups: the control group

($F(1, 145) = 1196.29$, $p<0.001$, partial $\eta^2 = 0.892$), the aMCI group ($F(1, 25) = 126.14$, $p<0.001$, partial $\eta^2 = 0.835$), the naMCI group ($F(1, 8) = 57.08$, $p<0.001$, partial $\eta^2 = 0.877$), the >MCI group ($F(1, 5) = 15.73$, $p = 0.01$, partial $\eta^2 = 0.759$) and the low mood group ($F(1, 9) = 46.64$, $p<0.001$, partial $\eta^2 = 0.838$).

There was a statistically significant interaction between classification and task complexity on mean reaction time ($F(4, 192) = 3.436$, $p = 0.01$, partial $\eta^2 = .067$), indicating that performance disproportionately slowed with increasing task complexity for some groups compared to others (see Figure 5.3). Looking at the difference between tasks in group mean RTs, compared to the controls, persons with low mood demonstrated the largest slowing from the simple to the complex task, followed by persons with naMCI and >MCI. Conversely, persons with aMCI demonstrated less slowing from the simple to the complex task (see Table 5.2).

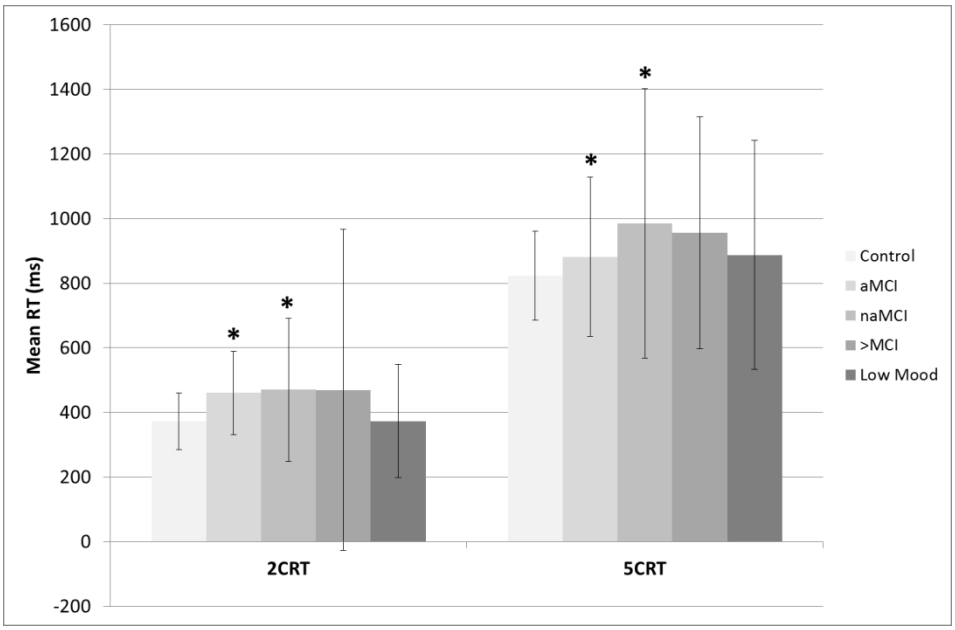


Figure 5.2: Mean RT for 2CRT and 5CRT task for each classification group

Mean reaction time (in ms) is presented for the 2CRT (left hand side) and 5CRT (right hand side) tasks. The RT is presented per classification group. The height of the bars indicates median group performance and the error bars indicate IQR; * indicates a difference from controls at $p \leq 0.01$ (analyses performed on transformed data for 146 controls, 26 aMCI, 9 naMCI, 6 >MCI, 10 low mood where data were available for both tasks).

Table 5.2: 2CRT task and TST task mean reaction times and reaction time variability by classification

		Control (n=164)	aMCI (n=28)	naMCI (n=11)	>MCI (n=9)	Low Mood (n = 12)
2CRT Task	**Mean RT (ms)**[†]	373.31 (87.56)	461.03 (128.94)*[b]	470.24 (221.43)*[c]	469.47 (496.38)	372.86 (175.57)
	RT IIV[†]	0.22 (0.09)	0.25 (0.14)	0.20 (0.10)	0.33 (0.10)*[c]	0.27 (0.07)
TST Task	**Mean RT**[‡]	822.79 (137.65)	881.66 (246.82)*[c]	984.71 (417.25)*[c]	956.67 (359.23)	887.61 (354.85)
	RT IIV[‡]	0.13 (0.06)	0.15 (0.11)	0.18 (0.22)	0.18 (0.19)	0.18 (0.13)
TST – 2CRT	**RT difference**	449.5	420.6	514.5	487.2	514.8

Values reported are median (IQR)

[†]n = 156 control, 26 aMCI; [‡]n = 154 control, 9 naMCI, 6 >MCI, 10 low mood

* $p \leq 0.01$ for difference from controls (analyses performed on transformed data for 146 controls, 26 aMCI, 9 naMCI, 6 >MCI, 10 low mood where data were available for both tasks).

[b]Cohen's d = 0.5 – 0.8 (medium effect size); [c]Cohen's d > 0.8 (large effect size)

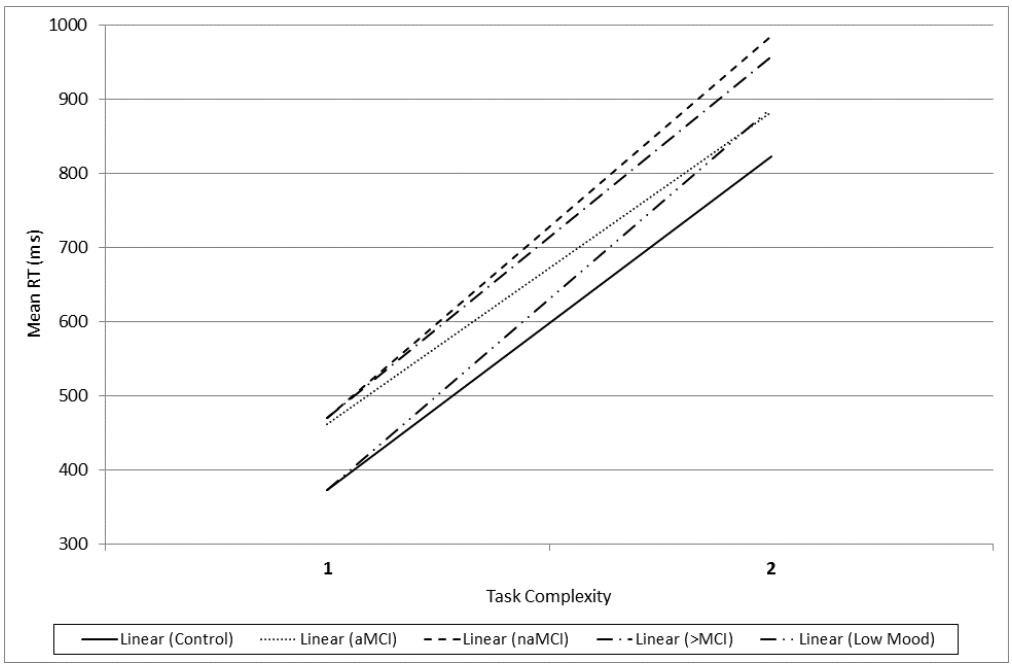


Figure 5.3: Mean RT x Task Complexity interaction

Mean reaction time (in ms) vs. task complexity (where 1 = 2CRT and 2 = 5CRT) plotted for each classification group (values plotted are medians)

Intraindividual Variability in Reaction Time

Table 5.2 and Figure 5.4 present the RT IIV of the participants by classification for each task. Results from the one way ANOVAs suggested that there were statistically significant differences between groups for RT IIV on the 2CRT task ($F(4, 209) = 4.277$, $p = 0.002$, partial $\eta^2 = 0.076$) and on the 5CRT task ($F(4, 202) = 4.224$, $p = 0.003$, partial $\eta^2 = 0.077$). Post hoc analysis revealed that participants with cognitive difficulties beyond MCI (>MCI) were significantly more variable than controls on the 2CRT task ($p = 0.02$) and the effect size for this comparison was large (Cohen's $d = 1.02$). However, for the 5CRT task, post-hoc analysis did not reveal any significant pairwise comparisons.

All groups were more variable on the simple task (2CRT) compared to the more complex task (5CRT). This difference was only statistically significant for: the control

group ($F(1, 145) = 133.54$, $p<0.001$, partial $\eta^2 = 0.479$), the aMCI group ($F(1, 25) = 15.61$, $p = 0.001$, partial $\eta^2 = 0.384$) and the low mood group ($F(1, 9) = 5.90$, $p = 0.038$, partial $\eta^2 = 0.396$).

There was no statistically significant interaction between classification and task complexity on RT IIV.

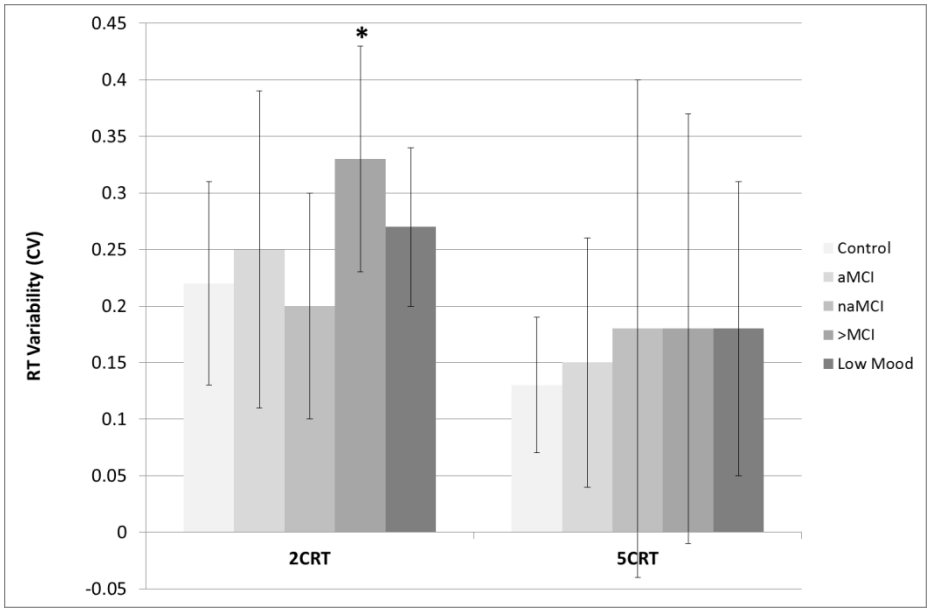


Figure 5.4: RT variability (CV) for 2CRT and 5CRT task for each classification group

Reaction time variability (measured as the coefficient of variation, CV) is presented for the 2CRT (left hand side) and 5CRT (right hand side) tasks. The RT variability is presented per classification group. The height of the bars indicates median group performance and the error bars indicate IQR; * indicates a difference from controls at $p \leq 0.01$ (analyses performed on transformed data for 146 controls, 26 aMCI, 9 naMCI, 6 >MCI, 10 low mood where data were available for both tasks).

Accuracy

2CRT Incorrect Responses

The numbers of incorrect responses made on the 2CRT task by participants by classification are shown in Table 5.3. The numbers of incorrect responses were generally low across all participants. There were no statistically significant differences

in the number of 2CRT incorrect responses between the different classifications, $\chi^2(4) = 5.421$, p=0.247.

5CRT Misses

The numbers of misses made on the 5CRT task by participants by classification are shown in Table 5.3. Again, the numbers of misses were generally low across all participants. There were no statistically significant differences in the number of 5CRT misses between the different classifications, $\chi^2(4) = 9.453$, p=0.051.

5CRT Spatial Accuracy

Table 5.3 presents the mean spatial accuracy and spatial accuracy IIV on the 5CRT task of the participants by classification. There were no statistically significant differences in 5CRT mean spatial accuracy between the different classifications, $F(4, 202) = 1.393$, p = 0.238. Spatial accuracy IIV on the 5CRT task revealed a marginally statistically significant difference between the classifications, $F(4, 202) = 2.544$, p = 0.04, partial η^2 = 0.048. Post-hoc analyses, however, didn't reveal any significant pairwise comparisons.

Table 5.3: 2CRT task and TST task accuracy scores by classification

		Control (n=164)	aMCI (n=28)	naMCI (n=11)	>MCI (n=9)	Low Mood (n = 12)
2CRT Task	**Number of errors**	0 (2)	1 (3)	0 (1)	2 (4)	0 (3)
	Number of misses	1 (3)	1 (2)	1 (2)	3 (5)	2.5 (3)
TST Task	**Mean Spatial Accuracy (pixels)**	23.23 (7.38)	21.28 (8.30)	20.34 (4.25)	25.77 (13.40)	23.83 (8.98)
	Spatial Accuracy IIV	0.55 (0.12)	0.57 (0.13)	0.58 (0.21)	0.63 (0.25)	0.65 (0.19)

Values reported are median (IQR)

[†]n = 156 control, 26 aMCI; [‡]n = 154 control, 9 naMCI, 6 >MCI, 10 low mood

Predictive Ability of Scores

Multinomial regressions were performed to investigate the predictive abilities of the RT task derived measures. Regressions were performed for those measures that had demonstrated statistically significant differences between classifications: (1) Mean RT on the 2CRT task, (2) Mean RT on the 5CRT task, (3) RT IIV on the 2CRT task, (4) RT IIV on the 5CRT task and (5) spatial accuracy IIV on the 5CRT task. Regressions were also performed for the two brief cognitive tests that were investigated in Chapter 4 ((6) M@T and (7) TYM), in order that direct comparisons could be made. In order to avoid any confounding influence of age and NART IQ, both of which have been shown to be associated with increased risk of cognitive decline and dementia (Chen, Lin, & Chen, 2009; Pavlik, Doody, Massman, & Chan, 2006), they were included as covariates in each regression.

Table 5.4 summarises the results of the multinomial regression analyses. The overall hit ratio (% of cases with classifications correctly predicted) of the obtained models ranged from 72.3% to 78.3%. Of the RT measures, the best model was obtained using 5CRT mean RT (overall hit ratio = 76.7%); however, out of all measures, the best model was obtained using M@T (overall hit ratio = 78.3%). The groups for which a measure was found to be significantly predictive (in comparison with controls) have been highlighted in Table 5.4. For both the 2CRT and 5CRT tasks, the mean RT measure was found to significantly predict both aMCI and naMCI groups. However, the RT IIV measure was only significantly predictive of the aMCI group and only for the 2CRT task.

Interestingly, the 5CRT spatial accuracy IIV measure was found to significantly predict low mood and the M@T and TYM significantly predicted all classifications. Notably, in eleven out of the twenty tested group comparisons for the RT task derived measures

only, age and/or NART IQ were significant predictors of group classification. All models were deemed to be a good fit according to a comparison of the final vs. intercept only models and almost all (except the M@T model) according to the Pearson statistic.

Table 5.4: Multinomial regression analyses discriminating classifications (vs. control) for (1) 2CRT mean RT, (2) 2CRT RT IIV, (3) 5CRT mean RT, (4) 5CRT RT IIV, (5) 5CRT spatial accuracy IIV; (6) M@T, (7) TYM

(1) 2CRT mean RT: $X^2(4) = 10.915$, p = 0.03 (OHR = 73.7%)

Goodness of Fit: Pearson, $X^2(836) = 773.499$, p = 0.940; Final vs. Intercept only, $X^2(12)=36.1$, p<0.001

vs. Control	**B**	**OR**	**95% CI**
aMCI	0.004**	1.004	1.001 – 1.007
naMCI	0.004*	1.004	1.000 – 1.008
>MCI[a]	0.004	1.004	1.000 – 1.007
Low Mood[b]	0.001	1.001	0.996 – 1.006

(2) 2CRT RT IIV[$]: $X^2(4) = 7.204$, p = 0.126 (OHR = 72.3%)

Goodness of Fit: Pearson, $X^2(836) = 762.83$, p = 0.97; Final vs. Intercept only, $X^2(12)=32.39$, p=0.001

vs. Control	**B**	**OR**	**95% CI**
aMCI	0.460*	1.584	1.028 – 2.441
naMCI	-0.207	0.813	0.366 – 1.808
>MCI[a]	0.513	1.670	0.944 – 2.957
Low Mood[b]	0.351	1.420	0.785- 2.569

(3) 5CRT mean RT: $X^2(4) = 22.355$, p < 0.001 (OHR = 76.7%)

Goodness of Fit: Pearson, $X^2(808) = 689.74$, p = 0.999; Final vs. Intercept only, $X^2(12)=51.71$, p<0.001

vs. Control	**B**	**OR**	**95% CI**
aMCI	0.005**	1.005	1.003 – 1.008
naMCI	0.006**	1.006	1.002 – 1.009
>MCI[a]	0.003	1.003	0.999 – 1.008
Low Mood	0.003	1.003	0.999 – 1.007

(4) 5CRT RT IIV[$]: $X^2(4) = 5.988$, p = 0.200 (OHR = 74.8%)

Goodness of Fit: Pearson, $X^2(808) = 677.58$, p = 1.000; Final vs. Intercept only, $X^2(12)=35.35$, p<0.001

vs. Control	**B**	**OR**	**95% CI**
aMCI[a,b]	0.267	1.306	0.917 – 1.861
naMCI[b]	0.413	1.511	0.908 – 2.514
>MCI[a]	0.537	1.712	0.826 – 3.548
Low Mood	0.386	1.471	0.924 – 2.343

(5) 5CRT Spatial Accuracy IIV[$]: $X^2(4) = 5.512$, p = 0.239 (OHR = 74.3%)

Goodness of Fit (Pearson): $X^2(808) = 770.09$, p = 0.83; Final vs. Intercept only, $X^2(12)=34.87$, p<0.001

vs. Control	**B**	**OR**	**95% CI**
aMCI[a,b]	0.019	1.019	0.745 – 1.394
naMCI[b]	0.315	1.370	0.934 – 2.010

>MCI[a]	0.188	1.206	0.739 – 1.969
Low Mood	0.356*	1.428	1.017 – 2.004

(6) M@T: $X^2(4) = 80.745$, p < 0.001 (OHR = 78.3%)

Goodness of Fit: Pearson, $X^2(856) = 1206.7$, p <0.001; Final vs. Intercept only, $X^2(12)=105.7$, p<0.001

vs. Control	**B**	**OR**	**95% CI**
aMCI	-0.347**	0.707	0.634 – 0.788
naMCI	-0.203**	0.816	0.721 – 0.924
>MCI	-0.389**	0.677	0.586 – 0.782
Low Mood	-0.152*	0.859	0.757 – 0.974

(7) TYM: $X^2(4) = 60.369$, p < 0.001 (OHR = 78.0%)

Goodness of Fit (Pearson): $X^2(848) = 815.77$, p = 0.78; Final vs. Intercept only, $X^2(12)=88.78$, p<0.001

vs. Control	**B**	**OR**	**95% CI**
aMCI	-0.365**	0.694	0.609 – 0.792
naMCI	-0.176*	0.839	0.708 – 0.994
>MCI	-0.519**	0.595	0.487 – 0.727
Low Mood	-0.239**	0.787	0.672 – 0.922

Notes: Age and NART IQ were included as covariates throughout all analyses and were non-significant predictors except where indicated ([a] Age significant predictor at p<0.05;[b] NART IQ significant predictor at p<0.05)

[$]Variability measures were multiplied by a factor of 10 prior to running multinomial regressions

Discussion

This study aimed to investigate the effectiveness of measures derived from RT tasks for identifying people with cognitive impairment. Participants with aMCI, naMCI, cognitive difficulties beyond MCI (>MCI) and low mood, as well as control participants, were administered a simple (2CRT) and a more complex (5CRT) task. Mean RT, RT IIV and the number of errors/misses made were measured for both tasks. In addition, mean spatial accuracy and spatial accuracy IIV were measured for the more complex (5CRT) task. The discriminative ability of these variables was evaluated.

Mean RT & RT Variability

The results indicate that people with MCI (amnestic and non-amnestic) are significantly slower than healthy controls on simple and more complex RT tasks. These results reflect previous studies which have shown increased RTs in people with "mild cognitive disorders" (Anstey et al., 2007), aMCI (Gorus et al., 2008) and MCI (Christensen et al., 2005; Dixon et al., 2007) compared with normal controls. The results are similar to those reported by Gorus and colleagues (Gorus et al., 2008), which demonstrated increasing RT with increasing cognitive impairment severity, with the >MCI group being slower on average than the aMCI group, who were in turn slower than the control group on both tasks. This finding supports the proposition that age-related cognitive decline is associated with a decline in basic processing speed (Salthouse, 1996). Perhaps due to the small sample size and heterogeneity of the group, the difference between the >MCI group and the controls did not reach statistical significance. The heterogeneity of the group is reflected in the large variation in their mean RT scores (as measured by the inter-quartile range). It is also interesting to point out that this group in fact did not score significantly lower than the MCI groups on any of the neuropsychological tests, apart from on the BADLS (which measures activities of daily living performance). Therefore, it is likely that the >MCI group were not as severely impaired as the mild-moderate Alzheimer's disease group included in the study by Gorus et al (2008).

In terms of RT variability, only the >MCI group demonstrated a significantly greater RT IIV score than the control group, and this was found in the simple RT task only. This reflects previous research which has demonstrated increased IIV in mild dementia populations (Burton et al., 2006; Hultsch et al., 2000) and supports the idea that IIV may be a behavioural indicator of compromised neurological mechanisms (Hultsch et

al., 2008; MacDonald et al., 2006). The results again approximately mirrored those by Gorus et al (Gorus et al., 2008), which demonstrated increasing RT IIV with increasing cognitive impairment severity, with the >MCI group being more variable than the aMCI group, who were in turn more variable on average than the control group on both tasks. However, perhaps due to the small sample sizes, the differences in IIV between the aMCI group and controls were not statistically significant.

To the author's knowledge, this is the first study to present these RT task derived measures for people with non-amnestic MCI. These are people who have impairment in cognitive domains other than memory. As mentioned previously, these participants were significantly slower than controls on both tasks. Interestingly, they were also slower than the aMCI group on both tasks, and performed with a similar RT to the >MCI group on the 2CRT task and were even slower on the 5CRT task. In terms of RT IIV, compared to the control and aMCI groups, the naMCI group were slightly less variable on the 2CRT task but more variable on the 5CRT task. The differences were not statistically significant, however their poorer performance on the more complex task is interesting to point out and may be reflective of their impairments in executive function and attention which may have impacted on their ability to perform this more cognitively demanding task.

The low mood group performed with a similar RT to controls on the simple task but were slower on the more complex task. The low mood group were also more variable than controls on both tasks, although, perhaps due to the small sample size, none of these differences reached statistical significance. In their study, Bunce et al (2008) found an association between depression scores and variability in performance on more cognitively demanding tasks. A task that demonstrated particularly strong associations

with both older age and higher depression was one that involved measuring the RTs of participants searching for targets on the basis of colour and shape (a conjunctive visual search task). It was proposed that this finding was in line with the suggestion that attentional resources tend to be reduced in people with depression and that those resources available tend to be deployed towards depression-related thoughts resulting in cognitive deficits, particularly in more effortful processing (Bunce et al., 2008; Hartlage, Alloy, Vazquez, & Dykman, 1993).

Accuracy

The number of errors made on the 2CRT task and the number of misses made on the 5CRT task were generally low across all groups; however, for both tasks, the participants with cognitive difficulties beyond MCI (>MCI) made the most errors and misses (although not statistically significantly different from any other group). In their study, Gorus et al (2008) also found no significant differences in number of errors between cognitively healthy people and those with aMCI; however, they did find that those with mild-moderate AD made significantly more errors. Again, a statistically significant result may not have been found here due to the small sample size and heterogeneity of the >MCI group.

To the author's knowledge, there has been no previous research addressing spatial accuracy in relation to cognitive impairment. In terms of mean spatial accuracy performance, the MCI groups appeared to perform with slightly better spatial accuracy than the control group; whereas the >MCI and low mood groups performed slightly worse. This does not indicate a clear pattern between mean spatial accuracy and cognitive impairment and makes the result difficult to interpret. Overall there appeared

to be a significant effect of group classification on spatial accuracy IIV, however, upon further consideration of pairwise differences, no significant differences were found. However, there did appear to be a tendency towards an increase in variability in spatial accuracy with cognitive impairment severity (with the >MCI group showing increased variability compared to the MCI groups, which in turn showed increased variability compared to the control group). It is worth noting that the low mood group performed on average with the greatest variability in spatial accuracy.

Effect of Task Complexity

There was an interaction between task complexity and classification for mean RT but not for RT variability; this same effect was demonstrated in a study by Christensen et al (Christensen et al., 2005). All groups performed more slowly on the more complex compared with the simple RT task; however, some groups slowed disproportionately between tasks compared with the control group. In particular, the naMCI and low mood groups demonstrated the greatest slowing when switching from the simple to the more complex task. Participants with cognitive difficulties beyond MCI also demonstrated more slowing than the control group (but to a lesser extent than the naMCI and low mood groups). This finding is in agreement with previous research which has shown that increasing task complexity is usually associated with poorer performance in non-healthy groups (Hultsch et al., 2000) and perhaps not surprising since more demanding tasks require more effortful processing. Conversely, however, the aMCI group demonstrated less slowing than controls when switching to the more complex task.

As stated above, there was no interaction between task complexity and classification on the RT IIV measure. However, it is interesting to note that all groups performed with

reduced variability on the more complex task compared to the simple task. Previous research has shown that IIV tends to increase on more complex tasks, particularly in more impaired samples, such as those with Alzheimer's disease (Burton et al., 2006). Whereas in less impaired samples, studies have demonstrated a reduction in IIV on more complex tasks, for example, studies by Anstey et al (2007) and Christensen et al (2005) report lower IIV values on a 2CRT task compared with a simple RT task for both normative and mild cognitive disorder samples. The findings from this study replicate these latter studies, which used similar community-based recruitment strategies and included samples with milder cognitive disorders, rather than established dementia cases.

Predictive Ability

Of the RT task derived measures assessed, mean RT on the more complex task (5CRT) produced the highest overall hit ratio (76.7%) to correctly identify all classifications. In particular, this variable was found to be a statistically significant predictor of aMCI and naMCI vs. control. This finding reflects previous studies which have demonstrated that more complex RT tasks tend to enhance discrimination between cognitive groups (Bielak et al., 2010; Gorus et al., 2008; Hultsch et al., 2000). For example, in their study, Bielak et al (2010) demonstrated that mean RT on a more complex 2CRT task provided better discrimination between cognitive groups than mean RT on a simple finger tapping task. This study has replicated these findings by demonstrating that an increase in choice level (2 vs. 5) provides enhanced prediction of cognitive groups. Bielak et al (2010) also demonstrated that further increasing task complexity, by including task switching (i.e. switching between responding to the shape or colour of a

figure), further enhanced discrimination between cognitive groups. This level of complexity was not investigated in this study.

The variability measures were slightly less successful at discriminating between all classification groups; however, the 2CRT RT IIV measure was found to be a statistically significant predictor of aMCI vs. control and in fact demonstrated a larger odds ratio than those found for the mean RT measures. This finding indicates that IIV may in fact be better at identifying people with aMCI in particular than mean RT, a finding which has been suggested in previous studies (Dixon et al., 2007; Gorus et al., 2008). However, it is unexpected to find that these measures were not statistically significantly predictive of more severe cognitive impairment (i.e. the >MCI group) but again this may be due to the heterogeneity of this sample, as well as it's small sample size in comparison with the other groups.

Spatial accuracy variability was not found to be statistically significantly predictive of the MCI or >MCI groups, however, interestingly, was found to be predictive of low mood status. This again could tie in with the idea that attentional resources tend to be reduced in those with depression resulting in cognitive deficits, particularly in more demanding tasks (Bunce et al., 2008; Hartlage et al., 1993).

Additionally, it is important to note that, although the RT derived measures were moderately useful at predicting cognitive classification, they were less accurate than the two brief cognitive tests evaluated in the previous chapter, the M@T and TYM. These brief cognitive tests (M@T and TYM) demonstrated overall hit ratios of ~78% and were statistically significantly predictive of all classifications. This indicates that tests of episodic memory and other cognitive domains are perhaps more useful in identifying

cognitive impairment than measures such as processing speed. However, the preliminary nature of this work must be noted. The sample sizes of the cognitively impaired groups were small, which may have accounted for the lack of statistically significant results. In fact, a post-hoc power calculation revealed that 62 aMCI and 372 control participants would have been required to detect a difference of 0.04 in 2CRT IIV (as was observed in the study by Anstey and colleagues (2007)) with a power of 90%. RT task derived measures do still present a useful, simple alternative to the traditionally used memory tests due to the fact that they are not influenced by administrator bias, education level or language ability and future studies with larger sample sizes should be considered to further investigate the potential role of RT derived measures in identifying cognitive impairment.

Conclusion

In summary, the results indicate that people with cognitive impairment have deficits in processing speed compared with healthy controls. Task complexity appears to have an effect on performance, particularly causing increased slowing in people with non-amnestic MCI, low mood and more severe cognitive impairment. Of the RT task derived measures, mean RT on the more complex task was the best overall predictor of cognitive status and RT variability on the simple task was shown to be particularly predictive of aMCI. However, none of the RT task derived measures were as accurate at predicting cognitive status as the two brief cognitive tests previously evaluated, the M@T and TYM.

Chapter 6: Discussion

This thesis aimed to investigate the effectiveness of a range of brief cognitive tests to be used for identifying people with MCI in an efficient and accurate manner. There has been increasing interest in identifying people with MCI since they are at an increased risk of developing dementia and could be ideal candidates on which to trial new dementia prevention interventions. However, MCI is largely unrecognised in primary care since its diagnosis depends on complex neuropsychological assessment methods not usually available in this setting. There is a need for a brief, sensitive and well-validated cognitive test to be used at initial patient contact to identify those most at risk of developing dementia. The previous chapters have presented a body of work that included: (a) the identification of candidate brief cognitive tests, (b) the recruitment of a cohort of older people on which to validate the chosen tests and (c) the investigation of the performance of these tests within the identified cohort. This final Chapter will collate the main findings of the work contained within the thesis and discuss the strengths and limitations of the work, as well as make suggestions for future directions.

Chapter 2 reported the findings from a systematic review that was conducted to identify the brief cognitive tests that have been used to detect people with the amnestic form of MCI (aMCI) and to evaluate the evidence for their validity. The focus of the review was aMCI since it is the commonest subtype (Roberts et al., 2012) and is associated with elevated rates of conversion to Alzheimer's disease (AD) (Petersen et al., 2001), which is the most common form of dementia (Ferri et al., 2005). Tests chosen for inclusion were those that were deemed to be quick (taking less than 15 minutes) and simple (not requiring equipment or specialist staff) to administer to ensure their suitability for use in busy primary care and research settings.

It was discovered that over 40 brief cognitive tests have been developed and tested for the identification of aMCI. Verbal learning tests that assess word recall were found to be particularly sensitive at detecting aMCI (for example, short delay recall on the AVLT (Zhao et al., 2012) and on the CERAD (Chandler et al., 2005; Karrasch et al., 2005; Woodard et al., 2005), as well as wordlist learning on the HVLT (Gonzalez-Palau et al., 2013; Schrijnemaekers et al., 2006)). Tasks such as these assess episodic memory, a feature known to be impaired in aMCI and early AD (Petersen et al., 1999) and thought to be the result of early pathological changes in the medial temporal lobe (Braak & Braak, 1998). It was argued, however, that tests that assess multiple cognitive domains provide the potential for a more comprehensive assessment, with the ability to detect non-memory cognitive impairment that is also frequently found in patients with aMCI, defined as multi-domain aMCI (Economou et al., 2007). The MoCA (Nasreddine et al., 2005), which involves the assessment of memory, semantic knowledge, language, visuospatial processing, attention, orientation and executive function, was identified as the most comprehensively investigated test, displaying high sensitivity for aMCI and high test-retest reliability. However, the MoCA can take up to 15 minutes to administer and another, quicker multi-task test named the M@T (Ravaglia et al., 2005), which takes just 5-10 minutes to administer and also exhibited very high sensitivity for aMCI, was highlighted as a potentially useful test.

A strength of this review over a similar previous review conducted by Lonie and colleagues (Lonie et al., 2009), is that an in depth quality appraisal of the included studies using a well-established tool (QUADAS-2, (Whiting et al., 2011)) was conducted. This appraisal revealed that a large proportion of included studies were judged to be at a high risk of unblinding the patient assessment process by comparing the performance of people with known aMCI from secondary care settings with that of

opportunistically recruited “controls” from the community. It has been reported that studies such as these may exaggerate diagnostic accuracy (Lijmer et al., 1999; Whiting et al., 2004). In addition, studies often did not use the same reference standard for all participants, with aMCI patients often undergoing more extensive assessments of cognition than the “controls” who were assumed to have no cognitive impairment, which was usually verified by a briefer set of tests. It was therefore argued that further validation studies, with improved design, were warranted.

The experimental work covered within the thesis was designed to address these limitations of previous studies. Chapter 3 described the work involved in identifying a cohort of older people from the community on which to investigate the validity of the chosen candidate tests. By inviting volunteers from the community to take part, without prior knowledge of their cognitive status, there was reduced risk of unblinding the assessment process. In addition to reducing the risk of bias in the validation process, the community based approach also provided useful information on recruitment rates of people with MCI to inform future studies applying similar recruitment techniques.

A prevalence of 16.5% of MCI within the assessed population was found, with aMCI being twice as prevalent as naMCI. A review by Ward and colleagues (2012) reported large variation in MCI prevalence across international studies, ranging from 3 to 42%, with the variability largely being attributed to a lack of consensus in MCI criteria and implementation (Ward et al., 2012). However, the current estimate is similar to that reported by other studies using similar recruitment techniques and criteria (Artero et al., 2006; Luck et al., 2007). As discussed in the Chapter, however, the current estimate was based only on those people who were assessed, which equated to 6.2% of the contacted population, and when taking into account all people who were approached, the rate of

MCI in fact dropped to 1.02%. It was argued that this more conservative estimate should be considered when designing future studies applying similar recruitment techniques.

At the outset of the study, the aim had been to recruit 200 people with aMCI, in order to maximise the precision with which the sensitivity of the tests under investigation could be reported. Only a quarter of this sample size target was actually achieved (n=52). This was largely due to the fact that the prevalence of aMCI within people reporting subjective memory impairment was lower than expected (12.5% compared with a predicted 30%, based on studies by Benito-Leon et al (2010) and Mitchell et al (2008)). In addition, the response rate to the flyers was much lower than anticipated. A recommendation was made that any future studies employing similar postal based recruitment strategies should consider re-contacting any initial non-responders as the rate of MCI may have in fact been higher within the delayed responder cohort, as demonstrated in a study by Miyamoto and colleagues (Miyamoto et al., 2009). Of those people who did respond at initial contact, a large proportion had to be excluded due to having no informant available to answer the informant-reported ADL scale. The use of alternative, performance-based ADL measures that do not require input from an informant for future studies was discussed; however, these can be unrepresentative of everyday functioning and are unsuited to large scale population-based aMCI case finding due to being time consuming and expensive to administer (Gold, 2012).

Chapter 4 covered the findings from the validation study of the M@T (Rami et al., 2007) and the TYM (Brown et al., 2009), which were identified from the literature as being potentially useful brief cognitive tests for identifying people with aMCI. Both tests were simple and quick to administer, confirming their suitability for use in busy

clinical and research settings. However, the M@T was found to perform with higher diagnostic test accuracy (DTA) than the TYM, with higher sensitivity (85% vs. 63%) and similar specificity (84% vs. 87%) and higher overall accuracy as demonstrated by the AUC values (0.91 vs. 0.80). The study found that, for both tests, the diagnostic test accuracy was not as high as had been previously reported (Custodio et al., 2014; Hanyu et al., 2011; Munoz-Neira et al., 2014; Rami et al., 2010; Rami et al., 2007; Szczesniak et al., 2013). The discrepancies were attributed to the fact that a community-based recruitment strategy was applied here, with the aMCI population assessed likely to be less impaired than those included in previous studies, which have mostly recruited from secondary care-based settings, such as memory clinics and neurology departments. The study therefore provided more conservative estimates of DTA, more likely to be generalizable to unselected populations.

In Chapter 5, the use of reaction time task derived measures in identifying cognitive impairment was explored. Reaction time (RT) tasks provide a simple way of assessing processing speed, which has been reported to demonstrate deficits early in the Alzheimer's disease process (Backman et al., 2005). As well as mean level of performance, which has been shown to increase with cognitive impairment (Anstey et al., 2007; Dixon et al., 2007), the consistency with which an individual performs across trials within a task has also been suggested as being an important indicator of cognitive functioning (Hultsch & MacDonald, 2004; Jensen, 1992). RT task measures present an attractive alternative to the more traditionally used memory tests such as M@T and TYM since they are not influenced by administrator bias, education level or language ability.

As expected, people with cognitive impairment were found to be slower than healthy controls on a simple two choice reaction time task (2CRT) and a more complex five choice reaction time task (5CRT). People with cognitive impairment were also found to be more variable than healthy controls on the simple 2CRT task; however, the difference only reached significance in participants with cognitive difficulties beyond MCI. Mean RT on the more complex 5CRT task was found to be the most accurate RT task derived measure at predicting cognitive status overall, reflecting previous studies which have demonstrated that more complex RT tasks tend to enhance discrimination between cognitive groups (Bielak et al., 2010; Gorus et al., 2008; Hultsch et al., 2000). For aMCI in particular, RT variability on the 2CRT task demonstrated slightly better discriminative abilities than mean performance. However, although these RT derived measures were moderately useful at predicting cognitive classification, they were less accurate than either the M@T or TYM, indicating that tests of episodic memory and other cognitive domains are perhaps more useful in identifying cognitive impairment than measures such as processing speed. However, the preliminary nature of the work was stressed and it was suggested that future studies using larger samples of cognitively impaired people should be conducted to further investigate the potential of RT task derived measures in identifying cognitive impairment.

Limitations & Future Work

It could be argued that a limitation with the experimental work is that the classifications of cognitive impairment were not clinically verified. Rather, an algorithmic approach to cognitive classification was applied. Although this differs from usual clinical practice, which would involve the incorporation of clinical and neurological examination, it did enable the criteria to be applied in a standardised and objective manner, thereby ensuring reliability of the classifications (Petersen et al., 2014). It would have been

useful to have completed longitudinal follow-up of the cohort to ascertain which individuals went onto develop dementia and were therefore exhibiting a true prodromal state of dementia, however, due to time limitations this could not be completed within the scope of this study. Cards were sent out to participants at completion of the study, with an invitation for those interested in taking part in future studies to return the card and so there is perhaps potential to perform a longer term follow-up of those who responded as part of a future study. It would be particularly useful to investigate how accurate the brief cognitive tests were at detecting future dementia converters.

Another limitation of the study, as discussed previously, is that the target sample size of 200 aMCI participants was not reached, which limited the precision with which the sensitivity of the memory tests could be reported. Future studies with larger cohorts of people with aMCI are required to provide more precise estimates of sensitivity for the M@T and TYM.

The M@T performed with high diagnostic test accuracy, albeit at a reduced level compared to that reported previously in the literature, largely as a result of the community-based recruitment strategy. It would be interesting to evaluate how it might perform relative to other well-validated tests such as the MoCA (Nasreddine et al., 2005) in a similar setting and head-to-head comparative studies are warranted in the future. The M@T also requires some adaptation to the wordlist for use in English-speaking populations so that more easily distinguishable words are included to reduce the risk of words being misheard. Therefore, it would be useful to use an adapted M@T in any future validation studies.

The current study only included English speaking people as the cognitive tests used to fulfil the Petersen criteria assessment have only been validated in English. However, this requirement to speak English will have led to the exclusion of a large number of people, particularly in the diverse area of Bradford where the research took place, which includes a large South Asian community, with many of its older generation unable to speak English to an adequate standard (Fuller, 2013). Therefore, it could be argued that the included cohort was not very representative of the local area. Further studies using translated versions of the cognitive tests are warranted so that future research in this area may be more inclusive.

Reaction time derived measures were investigated as a potential alternative to the traditionally used memory tests for identifying cognitive impairment. Measures such as these are not influenced by administrator bias and do not have particular education level or language requirements, which means they are more inclusive by nature. Unfortunately these measures were not as accurate at predicting cognitive status as the M@T or TYM. However, they did provide some promising results, which could be explored further. For instance, the effect of task complexity on the performance of people with naMCI in particular could be investigated further by including increased levels of task complexity that require even more effortful processing, to see if the discriminative abilities improve with increased task complexity.

Conclusion

This thesis has provided novel information on the validity of brief cognitive tests assessed within a cohort of volunteers recruited from the community, without prior knowledge of their cognitive status. This has resulted in more conservative estimates of the validity of the tests and provided useful estimates of the recruitment rates of people

with MCI, which should be considered when designing future studies applying similar recruitment techniques.

The current study has provided a prevalence estimate for MCI of 16.5%, based on an assessed sample of 472 people aged 70 years and above. However, future studies using a similar recruitment strategy should take account of the fact that this rate is dramatically reduced to 1.02% when taking account of all people approached.

Of the brief tests investigated, the M@T was found to be the most effective at identifying aMCI. It was quick and simple to administer and was found to perform with high sensitivity and specificity for identifying aMCI within the assessed cohort. RT task derived measures were not as accurate at predicting cognitive status as the M@T. However, they did demonstrate some promising discriminative abilities and should be further explored as potentially useful alternatives to the traditionally used memory tests, which may be influenced by administrator bias, education level and language ability.

References

Agrell, B., & Dehlin, O. (1998). The clock-drawing test. *Age Ageing, 41 Suppl 3*, iii41-45. doi:10.1093/ageing/afs149

Ahmed, S., de Jager, C., & Wilcock, G. (2012). A comparison of screening tools for the assessment of mild cognitive impairment: preliminary findings. *Neurocase, 18*(4), 336-351. doi:http://dx.doi.org/10.1080/13554794.2011.608365

Ahmed, S., Mitchell, J., Arnold, R., Dawson, K., Nestor, P. J., & Hodges, J. R. (2008). Memory complaints in mild cognitive impairment, worried well, and semantic dementia patients. *Alzheimer Dis Assoc Disord, 22*(3), 227-235. doi:10.1097/WAD.0b013e31816bbd27

Ahmed, S., Mitchell, J., Arnold, R., Nestor, P. J., & Hodges, J. R. (2008). Predicting rapid clinical progression in amnestic mild cognitive impairment. *Dementia & Geriatric Cognitive Disorders, 25*(2), 170-177. doi:http://dx.doi.org/10.1159/000113014

Ahn, H.-J., Chin, J., Park, A., Lee, B. H., Suh, M. K., Seo, S. W., & Na, D. L. (2010). Seoul Neuropsychological Screening Battery-dementia version (SNSB-D): a useful tool for assessing and monitoring cognitive impairments in dementia patients. *Journal of Korean Medical Science, 25*(7), 1071-1076. doi:http://dx.doi.org/10.3346/jkms.2010.25.7.1071

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement, 7*(3), 270-279. doi:10.1016/j.jalz.2011.03.008

Alegret, M., Boada-Rovira, M., Vinyes-Junque, G., Valero, S., Espinosa, A., Hernandez, I., . . . Tarraga, L. (2009). Detection of visuoperceptual deficits in preclinical and mild Alzheimer's disease. *Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society, 31*(7), 860-867. doi:http://dx.doi.org/10.1080/13803390802595568

Alladi, S., Arnold, R., Mitchell, J., Nestor, P. J., & Hodges, J. R. (2006). Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychol Med, 36*(4), 507-515. doi:10.1017/s0033291705006744

Anstey, K. J., Mack, H. A., Christensen, H., Li, S. C., Reglade-Meslin, C., Maller, J., . . . Sachdev, P. (2007). Corpus callosum size, reaction time speed and variability in mild cognitive disorders and in a normative sample. *Neuropsychologia, 45*(8), 1911-1920. doi:10.1016/j.neuropsychologia.2006.11.020

Artero, S., Petersen, R., Touchon, J., & Ritchie, K. (2006). Revised criteria for mild cognitive impairment: validation within a longitudinal population study. *Dementia & Geriatric Cognitive Disorders, 22*(5-6), 465-470. doi:10.1159/000096287

Artero, S., & Ritchie, K. (2003). The detection of mild cognitive impairment in the general practice setting. *Aging Ment Health, 7*(4), 251-258. doi:10.1080/1360786031000120741

Backman, L., Jones, S., Berger, A. K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology, 19*(4), 520-531. doi:10.1037/0894-4105.19.4.520

Barker, W. W., Luis, C. A., Kashuba, A., Luis, M., Harwood, D. G., Loewenstein, D., . . . Duara, R. (2002). Relative frequencies of Alzheimer disease, Lewy body,

vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Dis Assoc Disord, 16*(4), 203-212.

Benito-Leon, J., Mitchell, A. J., Vega, S., & Bermejo-Pareja, F. (2010). A population-based study of cognitive function in older people with subjective memory complaints. *Journal of Alzheimer's Disease, 22*(1), 159-170. doi:10.3233/jad-2010-100972

Bielak, A. A., Hultsch, D. F., Strauss, E., Macdonald, S. W., & Hunter, M. A. (2010). Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. *Neuropsychology, 24*(6), 731-741. doi:10.1037/a0019802

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, i*, 307-310.

Bosanquet, K., Bailey, D., Gilbody, S., Harden, M., Manea, L., Nutbrown, S., & McMillan, D. (2015). Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. *BMJ Open, 5*(12), e008913. doi:10.1136/bmjopen-2015-008913

Braak, H., & Braak, E. (1998). Evolution of neuronal changes in the course of Alzheimer's disease. *J Neural Transm Suppl, 53*, 127-140.

Brown, J., Pengas, G., Dawson, K., Brown, L. A., & Clatworthy, P. (2009). Self administered cognitive screening test (TYM) for detection of Alzheimer' s disease: cross sectional study. *BMJ, 338*, b2030.

Bucks, R. S., Ashworth, D. L., Wilcock, G. K., & Siegfried, K. (1996). Assessment of Activities of Daily Living in Dementia: Development of the Bristol Activities of Daily Living Scale. *Age and Ageing, 25*, 113-120.

Bunce, D., Anstey, K. J., Christensen, H., Dear, K., Wen, W., & Sachdev, P. (2007). White matter hyperintensities and within-person variability in community-dwelling adults aged 60-64 years. *Neuropsychologia, 45*(9), 2009-2015. doi:10.1016/j.neuropsychologia.2007.02.006

Bunce, D., Bielak, A. A., Cherbuin, N., Batterham, P. J., Wen, W., Sachdev, P., & Anstey, K. J. (2013). Utility of intraindividual reaction time variability to predict white matter hyperintensities: a potential assessment tool for clinical contexts? *J Int Neuropsychol Soc, 19*(9), 971-976. doi:10.1017/s1355617713000830

Bunce, D., Handley, R., & Gaines, S. O., Jr. (2008). Depression, anxiety, and within-person variability in adults aged 18 to 85 years. *Psychol Aging, 23*(4), 848-858. doi:10.1037/a0013678

Bunce, D. J., Warr, P. B., & Cochrane, T. (1993). Blocks in choice responding as a function of age and physical fitness. *Psychol Aging, 8*(1), 26-33.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*: Bury St Edmonds: Thames Valley Test Company.

Burton, C. L., Strauss, E., Hultsch, D. F., Moll, A., & Hunter, M. A. (2006). Intraindividual variability as a marker of neurological dysfunction: a comparison of Alzheimer's disease and Parkinson's disease. *J Clin Exp Neuropsychol, 28*(1), 67-83. doi:10.1080/13803390490918318

Buschke, H., Kuslansky, G., Katz, M., Stewart, W. F., Sliwinski, M. J., Eckholdt, H. M., & Lipton, R. B. (1999). Screening for dementia with the memory impairment screen. *Neurology, 52*(2), 231-238.

Cacho, J., Benito-Leon, J., Garcia-Garcia, R., Fernandez-Calvo, B., Vicente-Villardon, J. L., & Mitchell, A. J. (2010). Does the combination of the MMSE and clock drawing test (mini-clock) improve the detection of mild Alzheimer's disease and mild cognitive impairment? *Journal of Alzheimer's Disease, 22*(3), 889-896. doi:http://dx.doi.org/10.3233/JAD-2010-101182

Chandler, M. J., Lacritz, L. H., Hynan, L. S., Barnard, H. D., Allen, G., Deschner, M., . . . Cullum, C. M. (2005). A total score for the CERAD neuropsychological

battery. *Neurology, 65*(1), 102-106. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16009893

Chen, J. H., Lin, K. P., & Chen, Y. C. (2009). Risk factors for dementia. *J Formos Med Assoc, 108*(10), 754-764. doi:10.1016/s0929-6646(09)60402-2

Christensen, H., Dear, K. B., Anstey, K. J., Parslow, R. A., Sachdev, P., & Jorm, A. F. (2005). Within-occasion intraindividual variability and preclinical diagnostic status: is intraindividual variability an indicator of mild cognitive impairment? *Neuropsychology, 19*(3), 309-317. doi:10.1037/0894-4105.19.3.309

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Costa, A. S., Fimm, B., Friesen, P., Soundjock, H., Rottschy, C., Gross, T., . . . Reetz, K. (2012). Alternate-form reliability of the Montreal cognitive assessment screening test in a clinical setting. *Dementia & Geriatric Cognitive Disorders, 33*(6), 379-384. doi:http://dx.doi.org/10.1159/000340006

Crook, T., Bartus, R. T., Ferris, S. H., Whitehouse, P., Cohen, G. D., & Gershon, S. (1986). Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change - report of a national institute of mental health work group. *Developmental Neuropsychology, 2*(4), 261-276.

Custodio, N., Lira, D., Herrera-Perez, E., Nunez Del Prado, L., Parodi, J., Guevara-Silva, E., . . . Cortijo, P. (2014). The Memory Alteration Test Discriminates between Cognitively Healthy Status, Mild Cognitive Impairment and Alzheimer's Disease. *Dement Geriatr Cogn Dis Extra, 4*(2), 314-321. doi:10.1159/000365280

Deeks, J., Bossuyt, P., & Gatsonis, C. (2010). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.*

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test* (2nd ed.): Psychological Corporation, San Antonio, TX

World Health Organization and Alzheimer's Disease International. (2012). *Dementia: A Public Health Priority*. Geneva: World Health Organization.

Dierckx, E., Engelborghs, S., De Raedt, R., De Deyn, P. P., & Ponjaert-Kristoffersen, I. (2007). Differentiation between mild cognitive impairment, Alzheimer's disease and depression by means of cued recall. *Psychol Med, 37*(5), 747-755. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=17164030

Diniz, B. S., Nunes, P. V., Yassuda, M. S., Pereira, F. S., Flaks, M. K., Viola, L. F., . . . Forlenza, O. V. (2008). Mild cognitive impairment: cognitive screening or neuropsychological assessment? *Revista Brasileira de Psiquiatria, 30*(4), 316-321. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=19142405

http://www.scielo.br/pdf/rbp/v30n4/a03v30n4.pdf

Dixon, R. A., Garrett, D. D., Lentz, T. L., MacDonald, S. W., Strauss, E., & Hultsch, D. F. (2007). Neurocognitive markers of cognitive impairment: exploring the roles of speed and inconsistency. *Neuropsychology, 21*(3), 381-399. doi:10.1037/0894-4105.21.3.381

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., . . . Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol, 6*(8), 734-746. doi:10.1016/s1474-4422(07)70178-3

Dwamena, B. A., Sylvester, R., & Carlos, R. C. (2010). Midas: meta-analysis of diagnostic accuracy studies. http://fmwww.bc.edu/repec/bocode/m/midas.pdf Retrieved from http://fmwww.bc.edu/repec/bocode/m/midas.pdf

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Age differences in intra-individual variability in simple and choice reaction time: systematic review and meta-analysis. *PLoS One, 7*(10), e45759. doi:10.1371/journal.pone.0045759

Ebly, E. M., Hogan, D. B., & Parhad, I. M. (1995). Cognitive impairment in the nondemented elderly. Results from the Canadian Study of Health and Aging. *Arch Neurol, 52*(6), 612-619.

Ebly, E. M., Parhad, I. M., Hogan, D. B., & Fung, T. S. (1994). Prevalence and types of dementia in the very old: results from the Canadian Study of Health and Aging. *Neurology, 44*(9), 1593-1600.

Economou, A., Papageorgiou, S. G., Karageorgiou, C., & Vassilopoulos, D. (2007). Nonepisodic memory deficits in amnestic MCI. *Cogn Behav Neurol, 20*(2), 99-106. doi:10.1097/WNN.0b013e31804c6fe7

Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., . . . Scazufca, M. (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet, 366*(9503), 2112-2117. doi:10.1016/s0140-6736(05)67889-0

Freitas, S., Simoes, M. R., Alves, L., & Santana, I. (2013). Montreal cognitive assessment: Validation study for mild cognitive impairment and alzheimer disease. *Alzheimer Disease and Associated Disorders, 27*(1), 37-43. doi:http://dx.doi.org/10.1097/WAD.0b013e3182420bfe

Fujiwara, Y., Suzuki, H., Yasunaga, M., Sugiyama, M., Ijuin, M., Sakuma, N., . . . Shinkai, S. (2010). Brief screening tool for mild cognitive impairment in older Japanese: validation of the Japanese version of the Montreal Cognitive Assessment. *Geriatrics & gerontology international, 10*(3), 225-232. doi:http://dx.doi.org/10.1111/j.1447-0594.2010.00585.x

Fuller, K. (2013). *Understanding Bradford District (UBD): Ethnicity, language and religion* Retrieved from http://www.westyorkshireobservatory.org/

Ganguli, M., Dodge, H. H., Shen, C., & DeKosky, S. T. (2004). Mild cognitive impairment, amnestic type: an epidemiologic study. *Neurology, 63*(1), 115-121.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., . . . Winblad, B. (2006). Mild cognitive impairment. *Lancet, 367*(9518), 1262-1270. doi:10.1016/s0140-6736(06)68542-5

Geslani, D. M., Tierney, M. C., Herrmann, N., & Szalai, J. P. (2005). Mild cognitive impairment: an operational definition and its conversion rate to Alzheimer's disease. *Dementia & Geriatric Cognitive Disorders, 19*(5-6), 383-389. doi:10.1159/000084709

Gold, D. A. (2012). An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *J Clin Exp Neuropsychol, 34*(1), 11-34. doi:10.1080/13803395.2011.614598

Gonzalez-Palau, F., Franco, M., Jimenez, F., Parra, E., Bernate, M., & Solis, A. (2013). Clinical utility of the Hopkins Verbal Test-Revised for detecting Alzheimer's disease and mild cognitive impairment in Spanish population. *Archives of Clinical Neuropsychology, 28*(3), 245-253. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc7&AN=2013-14381-005

http://acn.oxfordjournals.org/content/28/3/245.full.pdf

Gorus, E., De Raedt, R., Lambert, M., Lemper, J. C., & Mets, T. (2008). Reaction times and performance variability in normal aging, mild cognitive impairment, and Alzheimer's disease. *Journal of Geriatric Psychiatry & Neurology, 21*(3), 204-218. doi:10.1177/0891988708320973

Grundman, M., Petersen, R. C., Ferris, S. H., Thomas, R. G., Aisen, P. S., Bennett, D. A., . . . Thal, L. J. (2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol, 61*(1), 59-66. doi:10.1001/archneur.61.1.59

Guo, Q.-h., Zhou, B., Zhao, Q.-h., Wang, B., & Hong, Z. (2012). Memory and Executive Screening (MES): a brief cognitive test for detecting mild cognitive impairment. *BMC Neurology, 12*, 119. doi:http://dx.doi.org/10.1186/1471-2377-12-119

Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform, 48*, 193-204. doi:10.1016/j.jbi.2014.02.013

Hanyu, H., Kume, K., Takada, Y., Onuma, T., & Iwamoto, T. (2009). The 1-minute Mental Status Examination in the memory clinic. *Journal of the American Geriatrics Society, 57*(6), 1130-1131. doi:http://dx.doi.org/10.1111/j.1532-5415.2009.02287.x

Hanyu, H., Maezono, M., Sakurai, H., Kume, K., Kanetaka, H., & Iwamoto, T. (2011). Japanese version of the Test Your Memory as a screening test in a Japanese memory clinic. *Psychiatry Res, 190*(1), 145-148. doi:http://dx.doi.org/10.1016/j.psychres.2011.04.025

Harbord, R. M., Deeks, J. J., Egger, M. S., Whiting, P., & Sterne, J. A. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics, 8*(2), 239-251.

Hartlage, S., Alloy, L. B., Vazquez, C., & Dykman, B. (1993). Automatic and effortful processing in depression. *Psychol Bull, 113*(2), 247-278.

Heilbronner, R. L., Sweet, J. J., Attix, D. K., Krull, K. R., Henry, G. K., & Hart, R. P. (2010). Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test administrations in clinical and forensic contexts. *Clin Neuropsychol, 24*(8), 1267-1278. doi:930077457 [pii]
10.1080/13854046.2010.526785 [doi]

Herbert, R. (2013). Confidence Interval Calculator. Retrieved from http://www.pedro.org.au/english/downloads/confidence-interval-calculator/

Hetherington, C. R., Stuss, D. T., & Finlayson, M. A. (1996). Reaction time and variability 5 and 10 years after traumatic brain injury. *Brain Inj, 10*(7), 473-486.

Howard, D., & Patterson, K. (1992). *Pyramids and palm trees: a test of semantic access from pictures and words*: Bury St Edmonds: Thames Valley Test Company.

Hughes, T. F., Snitz, B. E., & Ganguli, M. (2011). Should mild cognitive impairment be subtyped? *Curr Opin Psychiatry, 24*(3), 237-242. doi:10.1097/YCO.0b013e328344696b

Hultsch, D. F., & MacDonald, S. W. (2004). Intra-individual variability in performance as a theoretical window into cognitive aging. In R. A. Dixon, L. G. Nilsson, & L. Backman (Eds.), *New frontiers in cognitive aging* (pp. 65-88). Oxford, England: Oxford University Press.

Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *J Gerontol B Psychol Sci Soc Sci, 57*(2), P101-115.

Hultsch, D. F., MacDonald, S. W., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14*(4), 588-598.

Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. (2008). Intraindividual variability, cognition and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The*

handbook of aging and cognition (3rd ed., pp. 491-556). New York: Psychology Press.

Jack, C. R., Shiung, M. W., Weigand, S. D., & al, e. (2005). Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. *Neurology, 65*, 1227-1231.

Jackson, J. D., Balota, D. A., Duchek, J. M., & Head, D. (2012). White matter integrity and reaction time intraindividual variability in healthy aging and early-stage Alzheimer disease. *Neuropsychologia, 50*(3), 357-366. doi:10.1016/j.neuropsychologia.2011.11.024

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences, 13*(8), 869-881.

Jonker, C., Geerlings, M. I., & Schmand, B. (2000). Are memory complaints predictive for dementia? A review of clinical and population-based studies. *Int J Geriatr Psychiatry, 15*(11), 983-991.

Jungwirth, S., Fischer, P., Weissgram, S., Kirchmeyr, W., Bauer, P., & Tragl, K. H. (2004). Subjective memory complaints and objective memory impairment in the Vienna-Transdanube aging community. *J Am Geriatr Soc, 52*(2), 263-268.

Karrasch, M., Sinerva, E., Gronholm, P., Rinne, J., & Laine, M. (2005). CERAD test performances in amnestic mild cognitive impairment and Alzheimer's disease. *Acta Neurologica Scandinavica, 111*(3), 172-179. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=15691286

http://onlinelibrary.wiley.com/store/10.1111/j.1600-0404.2005.00380.x/asset/j.1600-0404.2005.00380.x.pdf?v=1&t=hpya3pxp&s=1e791b2426d523cc841bc8c735ba0cdab99ec15b

Kato, Y., Narumoto, J., Matsuoka, T., Okamura, A., Koumi, H., Kishikawa, Y., . . . Fukui, K. (2013). Diagnostic performance of a combination of mini-mental state examination and clock drawing test in detecting Alzheimer's disease. *Neuropsychiatric Disease and Treatment, 9*, 581-586. doi:http://dx.doi.org/10.2147/NDT.S42209

Knapp, M., & Prince, M. (2007). *Dementia UK - A report into the prevalence and cost of dementia prepared by the Personal Social Services Research Unit (PSSRU) at the London School of Economics and the Institute of Psychiatry at King's College London, for the Alzheimer's Society*. Retrieved from

Ladeira, R. B., Diniz, B. S., Nunes, P. V., & Forlenza, O. V. (2009). Combining cognitive screening tests for the evaluation of mild cognitive impairment in the elderly. *Clinics (Sao Paulo, Brazil), 64*(10), 967-973. doi:http://dx.doi.org/10.1590/S1807-59322009001000006

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol, 4*, 863. doi:10.3389/fpsyg.2013.00863

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174.

Lee, J.-Y., Dong Woo, L., Cho, S.-J., Na, D. L., Hong Jin, J., Kim, S.-K., . . . Maeng Je, C. (2008). Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry & Neurology, 21*(2), 104-110. doi:http://dx.doi.org/10.1177/0891988708316855

Leeflang, M. M., Moons, K. G., Reitsma, J. B., & Zwinderman, A. H. (2008). Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. *Clinical Chemistry, 54*(4), 729-737.

Lenehan, M. E., Klekociuk, S. Z., & Summers, M. J. (2012). Absence of a relationship between subjective memory complaint and objective memory impairment in mild cognitive impairment (MCI): is it time to abandon subjective memory complaint as an MCI diagnostic criterion? *Int Psychogeriatr, 24*(9), 1505-1514. doi:10.1017/s1041610212000695

Levy, R. (1994). Aging-associated cognitive decline. Working Party of the International Psychogeriatric Association in collaboration with the World Health Organization. *Int Psychogeriatr, 6*(1), 63-68.

Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bonsel, G. J., Prins, M. H., van der Meulen, J. H., & Bossuyt, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA, 282*(11), 1061-1066.

Department of Health (2009). *Living well with dementia: A National Dementia Strategy*.

Loewenstein, D. A., Acevedo, A., Agron, J., Ownby, R., Barker, W., Strauman, S., & Duara, R. (2009). The utility of a brief memory screen in the diagnosis of mild memory impairment in the elderly: preliminary study. *American Journal of Geriatric Psychiatry, 17*(5), 437-440. doi:http://dx.doi.org/10.1097/JGP.0b013e31819d3784

Lonie, J. A., Parra-Rodriguez, M. A., Tierney, K. M., Herrmann, L. L., Donaghey, C., O'Carroll, R. E., & Ebmeier, K. P. (2010). Predicting outcome in mild cognitive impairment: 4-year follow-up study. *Br J Psychiatry, 197*(2), 135-140. doi:10.1192/bjp.bp.110.077958

Lonie, J. A., Tierney, K. M., & Ebmeier, K. P. (2009). Screening for mild cognitive impairment: a systematic review. *Int J Geriatr Psychiatry, 24*(9), 902-915. doi:10.1002/gps.2208

Lovden, M., Li, S. C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: longitudinal data from the Berlin Aging Study. *Neuropsychologia, 45*(12), 2827-2838. doi:10.1016/j.neuropsychologia.2007.05.005

Luck, T., Luppa, M., Briel, S., & Riedel-Heller, S. G. (2010). Incidence of mild cognitive impairment: a systematic review. *Dementia & Geriatric Cognitive Disorders, 29*(2), 164-175. doi:10.1159/000272424

Luck, T., Riedel-Heller, S. G., Kaduszkiewicz, H., Bickel, H., Jessen, F., Pentzek, M., . . . Weyerer, S. (2007). Mild cognitive impairment in general practice: age-specific prevalence and correlate results from the German study on ageing, cognition and dementia in primary care patients (AgeCoDe). *Dementia & Geriatric Cognitive Disorders, 24*(4), 307-316. doi:10.1159/000108099

Luis, C. A., Keegan, A. P., & Mullan, M. (2009). Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *Int J Geriatr Psychiatry, 24*(2), 197-201. doi:http://dx.doi.org/10.1002/gps.2101

MacDonald, S. W., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: evidence from the Victoria Longitudinal Study. *Psychol Aging, 18*(3), 510-523. doi:10.1037/0882-7974.18.3.510

MacDonald, S. W., Li, S. C., & Backman, L. (2009). Neural underpinnings of within-person variability in cognitive functioning. *Psychol Aging, 24*(4), 792-808. doi:10.1037/a0017798

MacDonald, S. W., Nyberg, L., & Backman, L. (2006). Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. *Trends Neurosci, 29*(8), 474-480. doi:10.1016/j.tins.2006.06.011

Matthews, F. E., Stephan, B. C., McKeith, I. G., Bond, J., & Brayne, C. (2008). Two-year progression from mild cognitive impairment to dementia: to what extent do

different definitions agree? *J Am Geriatr Soc, 56*(8), 1424-1433. doi:10.1111/j.1532-5415.2008.01820.x

McDonald, A. M., Knight, R. C., Campbell, M. K., Entwistle, V. A., Grant, A. M., Cook, J. A., . . . Snowdon, C. (2006). What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials, 7*, 9. doi:10.1186/1745-6215-7-9

McLennan, S. N., Mathias, J. L., Brennan, L. C., & Stewart, S. (2011). Validity of the montreal cognitive assessment (MoCA) as a screening test for mild cognitive impairment (MCI) in a cardiovascular population. *Journal of Geriatric Psychiatry & Neurology, 24*(1), 33-38. doi:http://dx.doi.org/10.1177/0891988710390813

Mitchell, A. J. (2008a). The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis. *Int J Geriatr Psychiatry, 23*(11), 1191-1202. doi:10.1002/gps.2053

Mitchell, A. J. (2008b). Is it time to separate subjective cognitive complaints from the diagnosis of mild cognitive impairment? *Age Ageing, 37*(5), 497-499. doi:10.1093/ageing/afn147

Mitchell, A. J., & Shiri-Feshki, M. (2008). Temporal trends in the long term risk of progression of mild cognitive impairment: a pooled analysis. *J Neurol Neurosurg Psychiatry, 79*(12), 1386-1391. doi:10.1136/jnnp.2007.142679

Miyamoto, M., Kodama, C., Kinoshita, T., Yamashita, F., Hidaka, S., Mizukami, K., . . . Asada, T. (2009). Dementia and mild cognitive impairment among non-responders to a community survey. *Journal of Clinical Neuroscience, 16*(2), 270-276. doi:10.1016/j.jocn.2008.03.007

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6*(7), 1-6.

Molin, P., & Rockwood, K. (2016). The New Criteria for Alzheimer's Disease - Implications for Geriatricians. *Can Geriatr J, 19*(2), 66-73. doi:10.5770/cgj.19.207

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology, 43*(11), 2412-2414.

Muangpaisan, W., Intalapaporn, S., & Assantachai, P. (2010). Digit span and verbal fluency tests in patients with mild cognitive impairment and normal subjects in Thai-community. *Journal of the Medical Association of Thailand, 93*(2), 224-230. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20302005

Munoz-Neira, C., Henriquez Chaparro, F., Delgado, C., Brown, J., & Slachevsky, A. (2014). Test Your Memory-Spanish version (TYM-S): a validation study of a self-administered cognitive screening test. *Int J Geriatr Psychiatry, 29*(7), 730-740. doi:10.1002/gps.4055 [doi]

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc, 53*(4), 695-699. doi:JGS53221 [pii]
10.1111/j.1532-5415.2005.53221.x [doi]

Nelson, H. E., & Willison, J. (1991). *The National Adult Reading Test (NART)*: NFER-Nelson, Windsor, UK.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med, 17*(8), 857-872.

(2006). *Clinical guideline 42. Dementia: Supporting people with dementia and their carers in health and social care*.

National Institute for Health and Clinical Excellence. (2009). *Clinical guideline 90. Depression in adults: The treatment and management of depression in adults*.

Ozer, S., Young, J., Champ, C., & Burke, M. (2016). A systematic review of the diagnostic test accuracy of brief cognitive tests to detect amnestic mild cognitive impairment. *Int J Geriatr Psychiatry - UNDER REVIEW*.

Pavlik, V. N., Doody, R. S., Massman, P. J., & Chan, W. (2006). Influence of premorbid IQ and education on progression of Alzheimer's disease. *Dementia & Geriatric Cognitive Disorders, 22*(4), 367-377. doi:10.1159/000095640

Petersen, R. C. (2003). Conceptual Overview *Mild Cognitive Impairment: Aging to Alzheimer's Disease* (pp. 1-14). New York, NY: Oxford University Press.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine, 256*(3), 183-194. doi:10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: a concept in evolution. *J Intern Med, 275*(3), 214-228.

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., . . . Winblad, B. (2001). Current concepts in mild cognitive impairment. *Arch Neurol, 58*(12), 1985-1992.

Petersen, R. C., & Negash, S. (2008). Mild cognitive impairment: an overview. *CNS Spectr, 13*(1), 45-53.

Petersen, R. C., Parisi, J. E., Dickson, D. W., Johnson, K. A., Knopman, D. S., Boeve, B. F., . . . Kokmen, E. (2006). Neuropathologic features of amnestic mild cognitive impairment. *Arch Neurol, 63*(5), 665-672. doi:10.1001/archneur.63.5.665

Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., . . . Jack, C. R., Jr. (2009). Mild cognitive impairment: ten years later. *Arch Neurol, 66*(12), 1447-1455. doi:10.1001/archneurol.2009.266

Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., . . . Jack, C. R. (2009). Mild cognitive impairment: ten years later. *Arch Neurol, 66*(12), 1447-1455.

Petersen, R. C., Roberts, R. O., Knopman, D. S., Geda, Y. E., Cha, R. H., Pankratz, V. S., . . . Rocca, W. A. (2010). Prevalence of mild cognitive impairment is higher in men. The Mayo Clinic Study of Aging. *Neurology, 75*(10), 889-897. doi:10.1212/WNL.0b013e3181f11d85

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology, 56*(3), 303-308.

Psychology Research and Reference. (2016, 21 July). Retrieved from http://psychology.iresearchnet.com/social-psychology/social-psychology-research-methods/order-effects/

Rahman, T. T. A., & El Gaafary, M. M. (2009). Montreal Cognitive Assessment Arabic version: reliability and validity prevalence of mild cognitive impairment among elderly attending geriatric clubs in Cairo. *Geriatrics & gerontology international, 9*(1), 54-61. doi:http://dx.doi.org/10.1111/j.1447-0594.2008.00509.x

Rami, L., Bosch, B., Sanchez-Valle, R., & Molinuevo, J. L. (2010). The memory alteration test (M@T) discriminates between subjective memory complaints, mild cognitive impairment and Alzheimer's disease. *Arch Gerontol Geriatr, 50*(2), 171-174. doi:http://dx.doi.org/10.1016/j.archger.2009.03.005

Rami, L., Molinuevo, J. L., Sanchez-Valle, R., Bosch, B., & Villar, A. (2007). Screening for amnestic mild cognitive impairment and early Alzheimer's disease with M@T (Memory Alteration Test) in the primary care population. *Int J Geriatr Psychiatry, 22*(4), 294-304. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16998781

http://onlinelibrary.wiley.com/store/10.1002/gps.1672/asset/1672_ftp.pdf?v=1&t=hpzc4zzz&s=67de4c3cd7b83abf95379daae8255180aa55d2fd

Ravaglia, G., Forti, P., Maioli, F., Servadei, L., Martelli, M., Brunetti, N., . . . Mariani, E. (2005). Screening for mild cognitive impairment in elderly ambulatory patients with cognitive complaints. *Aging-Clinical & Experimental Research, 17*(5), 374-379. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16392411

Reisberg, B., Ferris, S. H., de Leon, M. J., & Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *Am J Psychiatry, 139*(9), 1136-1139. doi:10.1176/ajp.139.9.1136

Reisberg, B., Ferris, S. H., de Leon, M. J., Kluger, A., Franssen, E., Borenstein, J., & Alba, R. C. (1989). The stage specific temporal course of Alzheimer's disease: functional and behavioral concomitants based upon cross-sectional and longitudinal observation. *Prog Clin Biol Res, 317*, 23-41.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsycholgical Test Battery: Therapy and clinical interpretation*: Neuropsychological Press, Tucson, AZ.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J. P. M., Bossuyt, P., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specifcity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology, 58*, 982-990.

Ritchie, K. (2004). Mild cognitive impairment: an epidemiological perspective. *Dialogues Clin Neurosci, 6*(4), 401-408.

Roberts, R. O., Geda, Y. E., Knopman, D. S., Cha, R. H., Pankratz, V. S., Boeve, B. F., . . . Petersen, R. C. (2012). The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging. *Neurology, 78*(5), 342-351. doi:10.1212/WNL.0b013e3182452862

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychol Med, 44*(10), 2029-2040. doi:10.1017/s0033291713002535

Roth, M., Tym, E., Mountjoy, C. Q., Huppert, F. A., Hendrie, H., Verma, S., & Goddard, R. (1986). CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *Br J Psychiatry, 149*, 698-709.

Royal College of General Practitioners: RCGP Research Ready. (2016, July 16). Retrieved from http://www.rcgp.org.uk/researchready

Saka, E., Mihci, E., Topcuoglu, M. A., & Balkan, S. (2006). Enhanced cued recall has a high utility as a screening test in the diagnosis of Alzheimer's disease and mild cognitive impairment in Turkish people. *Archives of Clinical Neuropsychology, 21*(7), 745-751. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16979317

http://acn.oxfordjournals.org/content/21/7/745.full.pdf

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol Rev, 103*(3), 403-428.

Sarazin, M., Berr, C., De Rotrou, J., Fabrigoule, C., Pasquier, F., Legrain, S., . . . Dubois, B. (2007). Amnestic syndrome of the medial temporal type identifies prodromal AD: a longitudinal study.[Erratum appears in Neurology. 2008 May 20;70(21):2016]. *Neurology, 69*(19), 1859-1867. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=17984454

Scheurich, A., Muller, M. J., Siessmeier, T., Bartenstein, P., Schmidt, L. G., & Fellgiebel, A. (2005). Validating the DemTect with 18-fluoro-2-deoxy-glucose positron emission tomography as a sensitive neuropsychological screening test for early alzheimer disease in patients of a memory clinic. *Dementia & Geriatric Cognitive Disorders, 20*(5), 271-277. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16158009

Schneider, L. S., Mangialasche, F., Andreasen, N., Feldman, H., Giacobini, E., Jones, R., . . . Kivipelto, M. (2014). Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. *J Intern Med, 275*(3), 251-283. doi:10.1111/joim.12191

Schrijnemaekers, A. M. C., de Jager, C. A., Hogervorst, E., & Budge, M. M. (2006). Cases with mild cognitive impairment and Alzheimer's disease fail to benefit from repeated exposure to episodic memory tests as compared with controls. *Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society, 28*(3), 438-455. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16618630

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent Evidence and Development of a Shorter Version. *Clinical Gerontologist, 5*(1-2), 165-173.

Simon, J. R. (1968). Signal processing time as a function of aging. *J Exp Psychol, 78*(1), 76-80.

Smith, T., Gildeh, N., & Holmes, C. (2007). The Montreal Cognitive Assessment: validity and utility in a memory clinic setting. *Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie, 52*(5), 329-332. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=17542384

Stephan, B. C., & Brayne, C. (2014). Risk factors and screening methods for detecting dementia: a narrative review. *Journal of Alzheimer's Disease, 42 Suppl 4*, S329-338. doi:10.3233/jad-141413

Stephan, B. C., Matthews, F. E., McKeith, I. G., Bond, J., & Brayne, C. (2007). Early cognitive change in the general population: how do different definitions work? *J Am Geriatr Soc, 55*(10), 1534-1540. doi:10.1111/j.1532-5415.2007.01386.x

Stuss, D. T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: variability and consistency on reaction time tests. *Neuropsychology, 8*, 316-324.

Szczesniak, D., Wojtynska, R., & Rymaszewska, J. (2013). Test Your Memory (TYM) as a screening instrument in clinical practice - the Polish validation study. *Aging Ment Health, 17*(7), 863-868. doi:10.1080/13607863.2013.784957 [doi]

Takahashi, F., Awata, S., Sakuma, N., Inagaki, H., & Ijuin, M. (2012). Reliability and validity of A Quick Test of Cognitive Speed for detecting early-stage dementia in elderly Japanese. *Psychogeriatrics:The Official Journal of the Japanese Psychogeriatric Society, 12*(2), 75-82. doi:http://dx.doi.org/10.1111/j.1479-8301.2011.00388.x

Tombaugh, T. N. (2004). Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol, 19*(2), 203-214. doi:10.1016/s0887-6177(03)00039-8

Tsai, C. F., Lee, W. J., Wang, S. J., Shia, B. C., Nasreddine, Z., & Fuh, J. L. (2012). Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: Validation of the Taiwanese version of the MoCA and an item response theory analysis. *International Psychogeriatrics, 24*(4), 651-658. doi:http://dx.doi.org/10.1017/S1041610211002298

Vogel, A., Stokholm, J., Gade, A., Andersen, B. B., Hejl, A. M., & Waldemar, G. (2004). Awareness of deficits in mild cognitive impairment and Alzheimer's disease: do MCI patients have impaired insight? *Dementia & Geriatric Cognitive Disorders, 17*(3), 181-187. doi:10.1159/000076354

Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M., & Friedrich, F. (2006). The criterion validity of the Geriatric Depression Scale: a systematic review. *Acta Psychiatr Scand, 114*(6), 398-410. doi:10.1111/j.1600-0447.2006.00888.x

Ward, A., Arrighi, H. M., Michels, S., & Cedarbaum, J. M. (2012). Mild cognitive impairment: disparity of incidence and prevalence estimates. *Alzheimers Dement, 8*(1), 14-21. doi:10.1016/j.jalz.2011.01.002

Warrington, E. K. (1997). The Graded Naming Test: A Restandardisation. *Neuropsychol Rehabil, 7*, 143-146.

Warrington, E. K., & James, M. (1991). *The Visual Object and Space Perception Battery*: Pearson Assessment, London, UK.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain Cogn, 49*(3), 402-419.

Whiting, P., Rutjes, A. W., Reitsma, J. B., Glas, A. S., Bossuyt, P. M., & Kleijnen, J. (2004). Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med, 140*(3), 189-202.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . Bossuyt, P. M. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med, 155*(8), 529-536. doi:10.7326/0003-4819-155-8-201110180-00009

Whooley, M. A., Avins, A. L., Miranda, J., & Browner, W. S. (1997). Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med, 12*(7), 439-445.

Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association, 22*, 209-212.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., . . . Petersen, R. C. (2004). Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med, 256*(3), 240-246. doi:10.1111/j.1365-2796.2004.01380.x

Woodard, J. L., Dorsett, E. S. W., Cooper, J. G., Hermann, B. P., & Sager, M. A. (2005). Development of a brief cognitive screen for mild cognitive impairment and neurocognitive disorder. *Aging Neuropsychology and Cognition, 12*(4), 299-315. doi:10.1080/138255890968268

Yoshida, H., Terada, S., Honda, H., Kishimoto, Y., Takeda, N., Oshima, E., . . . Uchitomi, Y. (2012). Validation of the revised Addenbrooke's Cognitive Examination (ACE-R) for detecting mild cognitive impairment and dementia in a Japanese population. *Int Psychogeriatr, 24*(1), 28-37. doi:10.1017/S1041610211001190

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35.

Zhao, Q., Lv, Y., Zhou, Y., Hong, Z., & Guo, Q. (2012). Short-Term Delayed Recall of Auditory Verbal Learning Test Is Equivalent to Long-Term Delayed Recall for Identifying Amnestic Mild Cognitive Impairment. *PLoS ONE [Electronic Resource]*, *7*(12). doi:http://dx.doi.org/10.1371/journal.pone.0051157

Zhao, S., Guo, C., Wang, M., Chen, W., Wu, Y., Tang, W., & Zhao, Y. (2011). A clinical memory battery for screening for amnestic mild cognitive impairment in an elderly chinese population. *Journal of Clinical Neuroscience*, *18*(6), 774-779. doi:http://dx.doi.org/10.1016/j.jocn.2010.07.149

List of Appendices

Appendix 2.1 – CINAHL Search Strategy

☐	1	CINAHL “mild cognit* impair*”.ti,ab	1439	Apply Limits
☐	2	CINAHL MCI.ti,ab	959	Apply Limits
☐	3	CINAHL “early dementia”.ti,ab	125	Apply Limits
☐	4	CINAHL “early AD”.ti,ab	85	Apply Limits
☐	5	CINAHL “early Alzheimer*”.ti,ab	153	Apply Limits
☐	6	CINAHL “preclinical dementia”.ti,ab	18	Apply Limits
☐	7	CINAHL “pre-clinical dementia”.ti,ab	1	Apply Limits
☐	8	CINAHL “incipient dementia”.ti,ab	19	Apply Limits
☐	9	CINAHL “isolated memory impairment”.ti,ab	1	Apply Limits
☐	10	CINAHL (Amci OR Nmci OR Mmc).ti,ab	254	Apply Limits
☐	11	CINAHL (“N-MCI” OR “A-MCI” OR “M-MCI”).ti,ab	34	Apply Limits
☐	12	CINAHL “pre-clinical AD”.ti,ab	0	Apply Limits
☐	13	CINAHL “preclinical AD”.ti,ab	32	Apply Limits
☐	14	CINAHL “pre-clinical AAD”.ti,ab	0	Apply Limits
☐	15	CINAHL “pre-clinical Alzheimer*”.ti,ab	2	Apply Limits
☐	16	CINAHL AAMI.ti,ab	196	Apply Limits
☐	17	CINAHL ARCD.ti,ab	2	Apply Limits
☐	18	CINAHL CIND.ti,ab	57	Apply Limits

☐	19	CINAHL AACD.ti,ab	16	Apply Limits
☐	20	CINAHL SMC.ti,ab	119	Apply Limits
☐	21	CINAHL “questionable AD”.ti,ab	1	Apply Limits
☐	22	CINAHL “questionable dementia”.ti,ab	22	Apply Limits
☐	24	CINAHL “preclinical AD”.ti,ab	32	Apply Limits
☐	24	CINAHL “mild neurocognitive disorder*”.ti,ab	3	Apply Limits
☐	25	CINAHL “pre-clinical Alzheimer*”.ti,ab	2	Apply Limits
☐	26	CINAHL (“benign AND senescent AND forgetfulness).ti,ab	0	Apply Limits
☐	27	CINAHL BSF.ti,ab	12	Apply Limits
☐	28	CINAHL (“Limited cognitive disturbance” OR LCD).ti,ab	105	Apply Limits
☐	29	CINAHL (prodrom* adj2 dement*).ti,ab	24	Apply Limits
☐	30	CINAHL (“global deterioration scale” AND “stage 3”).ti,ab	1	Apply Limits
☐	31	CINAHL (“GDS 3” OR “stage 3 GDS”).ti,ab	3	Apply Limits
☐	32	CINAHL “Age associated memory impair*”.ti,ab	19	Apply Limits
☐	33	CINAHL “preclinical AAD”.ti,ab	0	Apply Limits
☐	34	CINAHL “preclinical Alzheimer*”.ti,ab	29	Apply Limits
☐	35	CINAHL “age related cognitive decline”.ti,ab	63	Apply Limits
☐	36	CINAHL ACMI.ti,ab	33	Apply Limits
☐	37	CINAHL “age associated cognitive decline”.ti,ab	9	Apply

					Limits
☐	38	CINAHL	“subjective memory complaint*”.ti,ab	80	Apply Limits
☐	39	CINAHL	“cognitive impairment-no* dement*”.ti,ab	44	Apply Limits
☐	40	CINAHL	(“minimal dementia” OR “MD”).ti,ab	5085	Apply Limits
☐	41	CINAHL	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33 OR 34 OR 35 OR 36 OR 37 OR 38 OR 39 OR 40	7958	Apply Limits
☐	41	CINAHL	(screen* adj3 (ora ge* OR method* OR scale* OR test* OR tool*)).ti,ab	9573	Apply Limits
☐	42	CINAHL	(screen* adj1 (examination* OR questionnaire*)).ti,ab	1039	Apply Limits
☐	43	CINAHL	((“clinical assessment tool*” OR “functional assessment*” OR “geriatric assessment*” OR “geriatric functional assessment*”)).ti,ab	2248	Apply Limits
☐	44	CINAHL	exp GERIATRIC ASSESSMENT/	9227	Apply Limits
☐	45	CINAHL	NEUROPSYCHOLOGICAL TESTS/	15173	Apply Limits
☐	46	CINAHL	PSYCHOMETRICS/	7161	Apply Limits
☐	47	CINAHL	(((cognit* OR memory OR mental) adj2 (assess* OR evaluat* OR screen* OR examin* OR diagnos* OR instrument* OR measure* OR questionnaire* OR scale* OR test*))).ti,ab	14019	Apply Limits
☐	48	CINAHL	41 OR 42 OR 43 OR 44 OR 45 OR 46 OR 47	50446	Apply Limits
☐	49	CINAHL	REPRODUCIBILITY OF RESULTS/	13319	Apply Limits
☐	50	CINAHL	VALIDATION STUDIES/	15038	Apply Limits
☐	51	CINAHL	ROC CURVE/	6330	Apply

					Limits
☐	52	CINAHL	SENSITIVITY AND SPECIFICITY/	25795	Apply Limits
☐	53	CINAHL	specific*.ti,ab	119750	Apply Limits
☐	54	CINAHL	ora ge*.ti,ab	41732	Apply Limits
☐	55	CINAHL	reliabil*.ti,ab	21576	Apply Limits
☐	56	CINAHL	accura*.ti,ab	36453	Apply Limits
☐	57	CINAHL	exp DIAGNOSTIC ERRORS/	8607	Apply Limits
☐	58	CINAHL	Roc.ti,ab	1480	Apply Limits
☐	58	CINAHL	49 OR 50 OR 51 OR 52 OR 53 OR 54 OR 55 OR 56 OR 57 OR 58	225045	Apply Limits
☐	145	CINAHL	41 AND 48 AND 58	474	Apply Limits
☐	146	CINAHL	145 [Limit to: Publication Year 1999-2013]	457	A

Appendix 2.2 – SRR Data Extraction Form

Reviewer:	Author(s):	Publication Year:

Study detail	
Brief statement of aims	
Authors' key conclusions	
Any design details/notes	
Duration of follow-up	

Participants	
Setting	
Country	
Characterisation	
Method of recruitment	
Inclusion criteria	
Exclusion criteria	

Clinical/neuropsychological assessment

Group 1	
Name	
Definition	

Group 2	
Name	
Definition	

Group 3	
Name	
Definition	

Participant Characteristics					
	Total	**Group 1**	**Group 2**	**Group 3**	**P value**
Number					
Age					
% Female					
Education, years					
MMSE					
Notes:					

Results – Diagnostic Test Accuracy				
Test Name				
Comparison groups				
AUC **Standard error** **P** **95% CI**				
Cut off				

Sensitivity %				
Specificity %				
PPV %				
NPV %				
Number of true positives				
Number of false negatives				
Number of false positives				
Number of true negatives				

Results – Test-Retest Reliability				
Test-retest reliability reported?		**YES**	**NO**	
Test Name				
Participants				
Time period				
Measure Used				
Result				

QUADAS-2 tool

Domain	Signalling Questions	Notes	YES	NO	UNCLEAR
1. Selection of patients	a. Was a case-control design avoided?				
	b. Was a consecutive or random sample of patients enrolled?				
2. Index test	c. Were the index test results interpreted without knowledge of the results of the reference standard?				
3. Reference standard	d. Were the reference standard results interpreted without knowledge of the results of the index test?				
4. Patient flow	e. Was there an appropriate interval between the index test and reference standard?				
	f. Did all patients receive the same reference standard?				
	g. Were all patients included in the analysis?				

Appendix 2.3: Forest plots summarising sensitivity and specificity for single task screens for identifying aMCI, with associated 95% confidence intervals

15-OT (12)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Alegret 2009	28	6	16	38	0.64 [0.48, 0.78]	0.86 [0.73, 0.95]

AQT-CF (72/73)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Takahashi 2012	21	6	4	19	0.84 [0.64, 0.95]	0.76 [0.55, 0.91]

CDT-CERAD (5)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	1	2	14	13	0.07 [0.00, 0.32]	0.87 [0.60, 0.98]

CDT-Command (8/9)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Cacho 2010	16	20	5	46	0.76 [0.53, 0.92]	0.70 [0.57, 0.80]
Kato 2013	30	6	30	43	0.50 [0.37, 0.63]	0.88 [0.75, 0.95]

CDT-Sunderland (ELD)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ladeira 2009	25	10	58	73	0.30 [0.21, 0.41]	0.88 [0.79, 0.94]

CDT-Sunderland (=<5)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ravaglia 2005	10	8	28	47	0.26 [0.13, 0.43]	0.85 [0.73, 0.94]

CDT-Wolf Klein (=<6)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ravaglia 2005	8	6	30	49	0.21 [0.10, 0.37]	0.89 [0.78, 0.96]

Constructional Praxis-Savings (60%)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	5	5	10	10	0.33 [0.12, 0.62]	0.67 [0.38, 0.88]

Digit Span Forward (12)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Muangpaisan 2010	49	9	28	21	0.64 [0.52, 0.74]	0.70 [0.51, 0.85]

Digit Span Backward (4)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Muangpaisan 2010	59	13	18	17	0.77 [0.66, 0.86]	0.57 [0.37, 0.75]

ECR-3rd Free Recall (9)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Saka 2006	10	7	8	26	0.56 [0.31, 0.78]	0.79 [0.61, 0.91]

ECR-Total Recall (42)

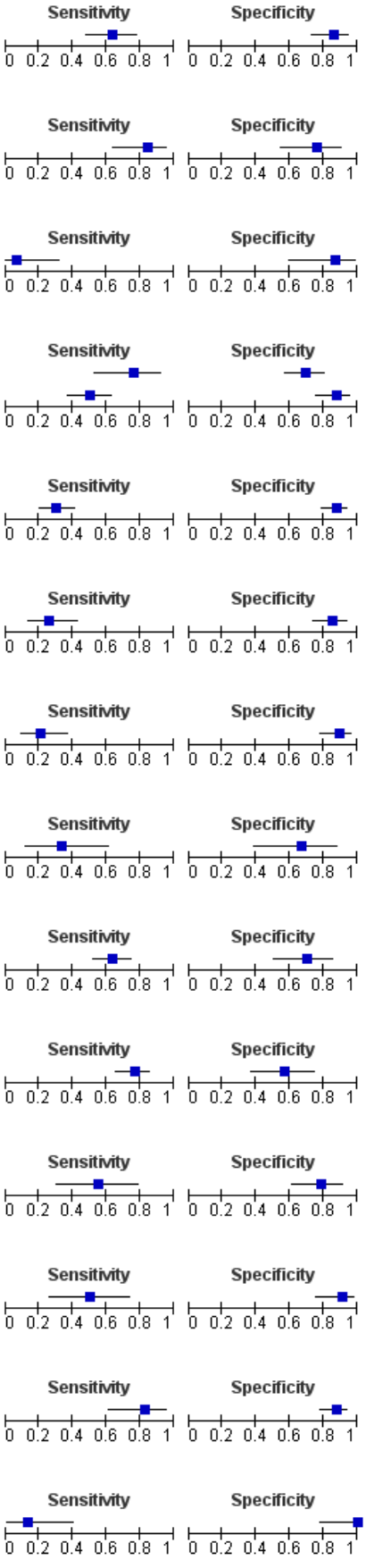
Study	TP	FP	FN	TN	Sensitivity	Specificity
Saka 2006	9	3	9	30	0.50 [0.26, 0.74]	0.91 [0.76, 0.98]

FBMS (7)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Loewenstein 2009	19	10	4	70	0.83 [0.61, 0.95]	0.88 [0.78, 0.94]

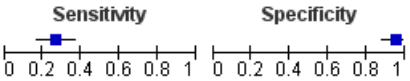
Naming-BNT-M (11)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	2	0	13	15	0.13 [0.02, 0.40]	1.00 [0.78, 1.00]



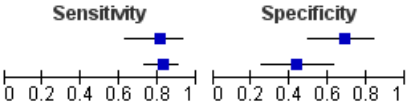
VFT-Animals (ELD)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ladeira 2009	22	4	61	79	0.27 [0.17, 0.37]	0.95 [0.88, 0.99]



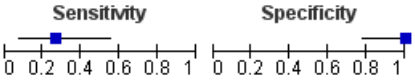
VFT-Animals (14)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Hanyu 2009	25	10	6	22	0.81 [0.63, 0.93]	0.69 [0.50, 0.84]
Muangpaisan 2010	64	17	13	13	0.83 [0.73, 0.91]	0.43 [0.25, 0.63]



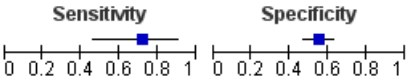
VFT-Animals (15)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	4	0	11	15	0.27 [0.08, 0.55]	1.00 [0.78, 1.00]



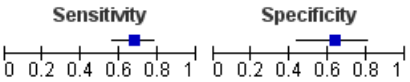
VFT-Animals (<20)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Woodard 2005	13	72	5	89	0.72 [0.47, 0.90]	0.55 [0.47, 0.63]



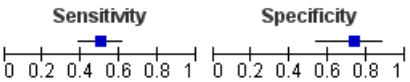
VFT-Fruits (15)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Muangpaisan 2010	52	11	25	19	0.68 [0.56, 0.78]	0.63 [0.44, 0.80]



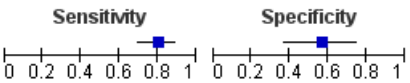
VFT-Letter Koh (9)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Muangpaisan 2010	39	8	39	22	0.50 [0.38, 0.62]	0.73 [0.54, 0.88]



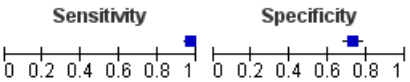
VFT-Letter Soh (7)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Muangpaisan 2010	62	13	15	17	0.81 [0.70, 0.89]	0.57 [0.37, 0.75]



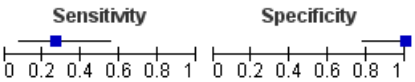
VLT-AVLT SR (=<2)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Zhao 2012	314	86	11	230	0.97 [0.94, 0.98]	0.73 [0.68, 0.78]



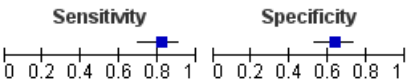
VLT-CERAD WLDR (6)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	4	0	11	15	0.27 [0.08, 0.55]	1.00 [0.78, 1.00]



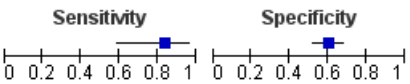
VLT-CERAD WLDR (6.5)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Chandler 2005	49	35	11	60	0.82 [0.70, 0.90]	0.63 [0.53, 0.73]



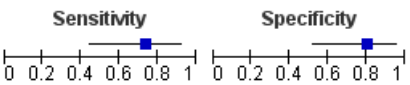
VLT-CERAD WLDR (<7)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Woodard 2005	15	64	3	97	0.83 [0.59, 0.96]	0.60 [0.52, 0.68]



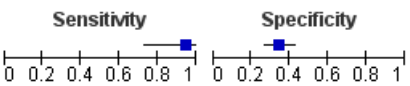
VLT-CERAD WLLE (20)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	11	3	4	12	0.73 [0.45, 0.92]	0.80 [0.52, 0.96]



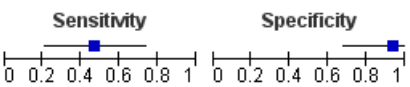
VLT-CERAD WLREDI (<10)

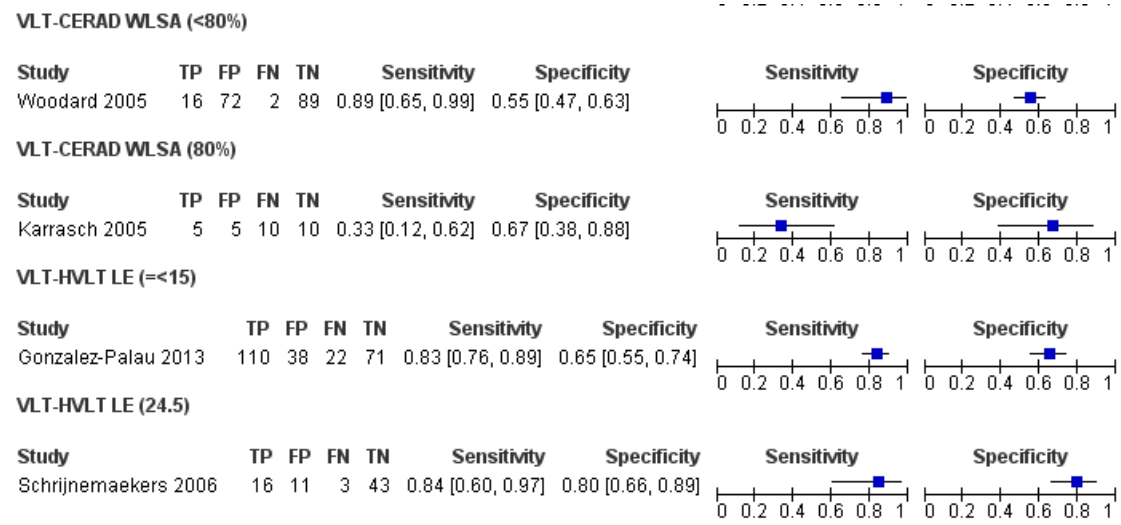
Study	TP	FP	FN	TN	Sensitivity	Specificity
Woodard 2005	17	105	1	56	0.94 [0.73, 1.00]	0.35 [0.27, 0.43]



VLT-CERAD WLRE (92%)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	7	1	8	14	0.47 [0.21, 0.73]	0.93 [0.68, 1.00]



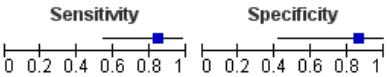


NB: cut-off score in brackets; for index test definitions, see Abbreviations

Appendix 2.4: Forest plots summarising sensitivity and specificity for multi-task screens for identifying aMCI, with associated 95% confidence intervals

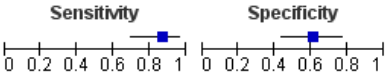
DemTect (=<13)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Scheurich 2005	11	1	2	6	0.85 [0.55, 0.98]	0.86 [0.42, 1.00]



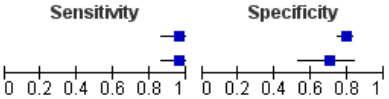
HDS-R (28/29)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Fujiwara 2010	26	14	4	22	0.87 [0.69, 0.96]	0.61 [0.43, 0.77]



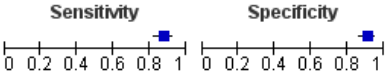
M@T (37)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Rami 2007	48	84	2	316	0.96 [0.86, 1.00]	0.79 [0.75, 0.83]
Rami 2010	48	11	2	26	0.96 [0.86, 1.00]	0.70 [0.53, 0.84]



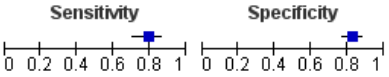
MES (=<72)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guo 2012	171	17	24	180	0.88 [0.82, 0.92]	0.91 [0.87, 0.95]



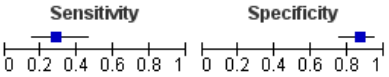
MES (=<75)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guo 2012	92	34	24	163	0.79 [0.71, 0.86]	0.83 [0.77, 0.88]



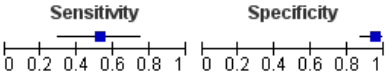
MMSE (<24)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ravaglia 2005	11	7	27	48	0.29 [0.15, 0.46]	0.87 [0.76, 0.95]



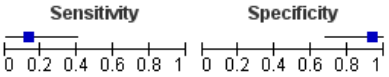
MMSE (24v25)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Cacho 2010	11	3	10	63	0.52 [0.30, 0.74]	0.95 [0.87, 0.99]



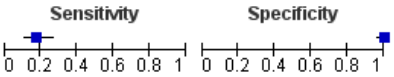
MMSE (25)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Karrasch 2005	2	1	13	14	0.13 [0.02, 0.40]	0.93 [0.68, 1.00]



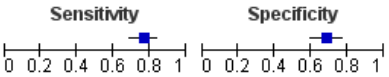
MMSE (<26)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Nasreddine 2005	17	0	77	90	0.18 [0.11, 0.27]	1.00 [0.96, 1.00]



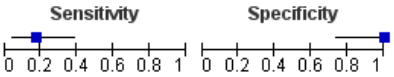
MMSE (=<26)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Gonzalez-Palau 2013	101	34	31	75	0.77 [0.68, 0.83]	0.69 [0.59, 0.77]



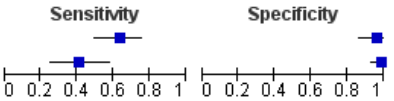
MMSE (26)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Smith 2007	4	0	19	12	0.17 [0.05, 0.39]	1.00 [0.74, 1.00]



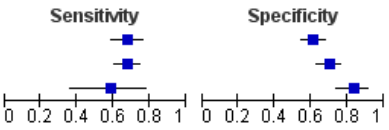
MMSE (26/27)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Kato 2013	38	2	22	47	0.63 [0.50, 0.75]	0.96 [0.86, 1.00]
Yoshida 2012	16	1	23	72	0.41 [0.26, 0.58]	0.99 [0.93, 1.00]



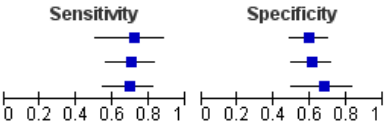
MMSE (=<27)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Guo 2012	79	76	37	121	0.68 [0.59, 0.76]	0.61 [0.54, 0.68]
Guo 2012	132	59	63	138	0.68 [0.61, 0.74]	0.70 [0.63, 0.76]
Luis 2009	14	12	10	62	0.58 [0.37, 0.78]	0.84 [0.73, 0.91]



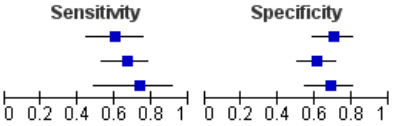
MMSE (27/28)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Diniz 2008	18	36	7	53	0.72 [0.51, 0.88]	0.60 [0.49, 0.70]
Diniz 2008	36	35	15	54	0.71 [0.56, 0.83]	0.61 [0.50, 0.71]
Hanyu 2011	32	11	14	23	0.70 [0.54, 0.82]	0.68 [0.49, 0.83]



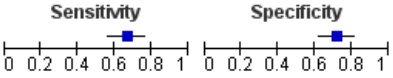
MMSE (28.5)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ahn 2010	26	23	17	54	0.60 [0.44, 0.75]	0.70 [0.59, 0.80]
Chandler 2005	40	37	20	58	0.67 [0.53, 0.78]	0.61 [0.51, 0.71]
Schrijnemaekers 2006	14	17	5	37	0.74 [0.49, 0.91]	0.69 [0.54, 0.80]



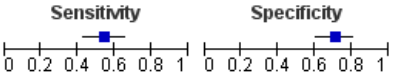
MMSE (<29)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Freitas 2013	60	25	30	65	0.67 [0.56, 0.76]	0.72 [0.62, 0.81]



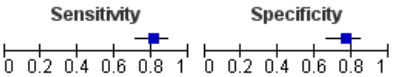
MMSE (ELD)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ladeira 2009	45	24	38	59	0.54 [0.43, 0.65]	0.71 [0.60, 0.81]



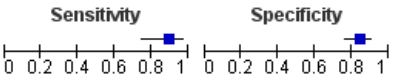
MoCA (<22)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Freitas 2013	73	21	17	69	0.81 [0.71, 0.89]	0.77 [0.67, 0.85]



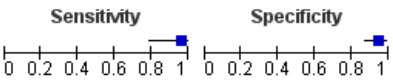
MoCA (22/23)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Lee 2008	33	18	4	97	0.89 [0.75, 0.97]	0.84 [0.76, 0.90]



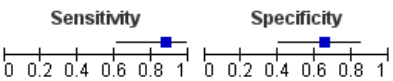
MoCA (23)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Luis 2009	23	4	1	70	0.96 [0.79, 1.00]	0.95 [0.87, 0.99]



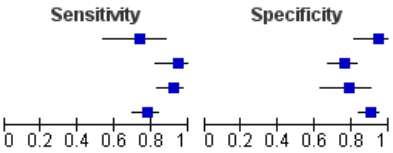
MoCA (23.5)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ahmed 2012	14	7	2	13	0.88 [0.62, 0.98]	0.65 [0.41, 0.85]



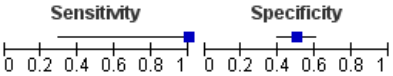
MoCA (23/24)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Fujiwara 2010	22	2	8	34	0.73 [0.54, 0.88]	0.94 [0.81, 0.99]
Lee 2008	35	28	2	87	0.95 [0.82, 0.99]	0.76 [0.67, 0.83]
Tsai 2012	65	8	6	30	0.92 [0.83, 0.97]	0.79 [0.63, 0.90]
Zhao 2011	116	15	34	135	0.77 [0.70, 0.84]	0.90 [0.84, 0.94]



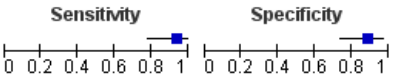
MoCA (<24)

Study	TP	FP	FN	TN	Sensitivity	Specificity
McLennan 2011	3	48	0	48	1.00 [0.29, 1.00]	0.50 [0.40, 0.60]



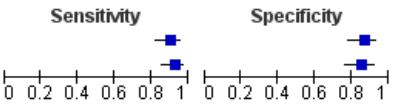
MoCA (25/26)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Fujiwara 2010	28	4	2	32	0.93 [0.78, 0.99]	0.89 [0.74, 0.97]



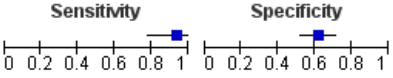
MoCA (<26)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Nasreddine 2005	84	12	9	78	0.90 [0.82, 0.95]	0.87 [0.78, 0.93]
Rahman 2009	87	13	7	77	0.93 [0.85, 0.97]	0.86 [0.77, 0.92]



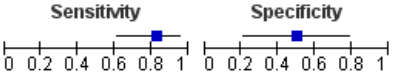
MoCA (=<26)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Costa 2012	28	38	2	62	0.93 [0.78, 0.99]	0.62 [0.52, 0.72]



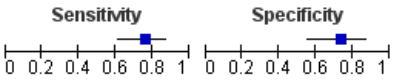
MoCA (26)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Smith 2007	19	6	4	6	0.83 [0.61, 0.95]	0.50 [0.21, 0.79]



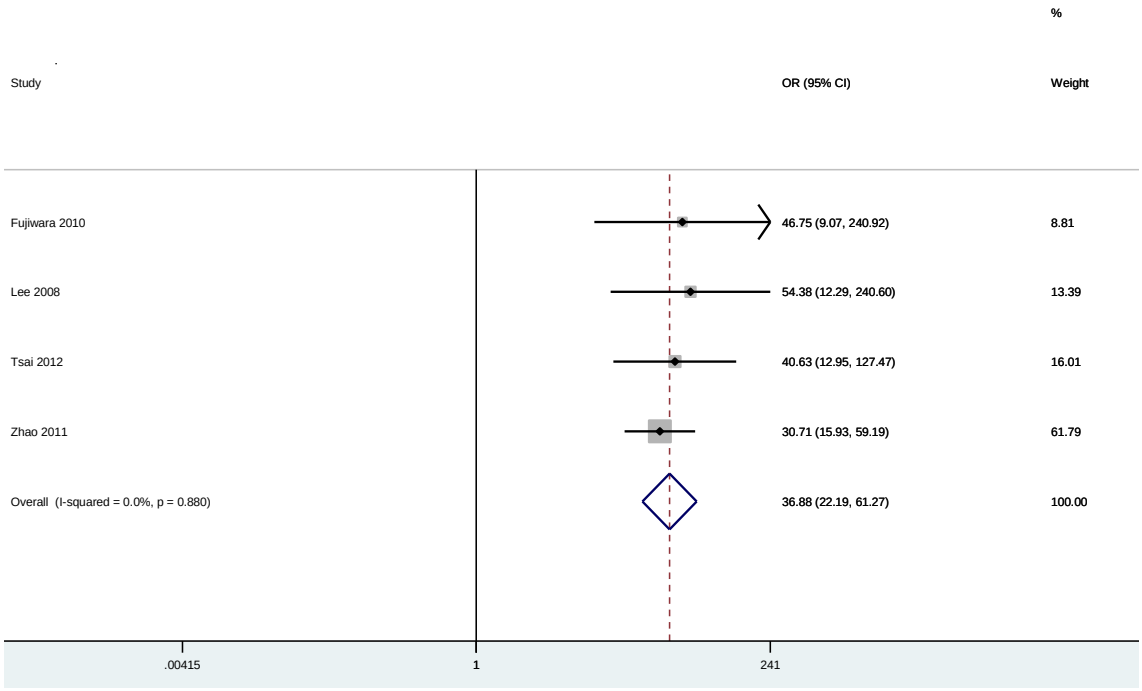
TYM (44/45)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Hanyu 2011	35	9	11	25	0.76 [0.61, 0.87]	0.74 [0.56, 0.87]

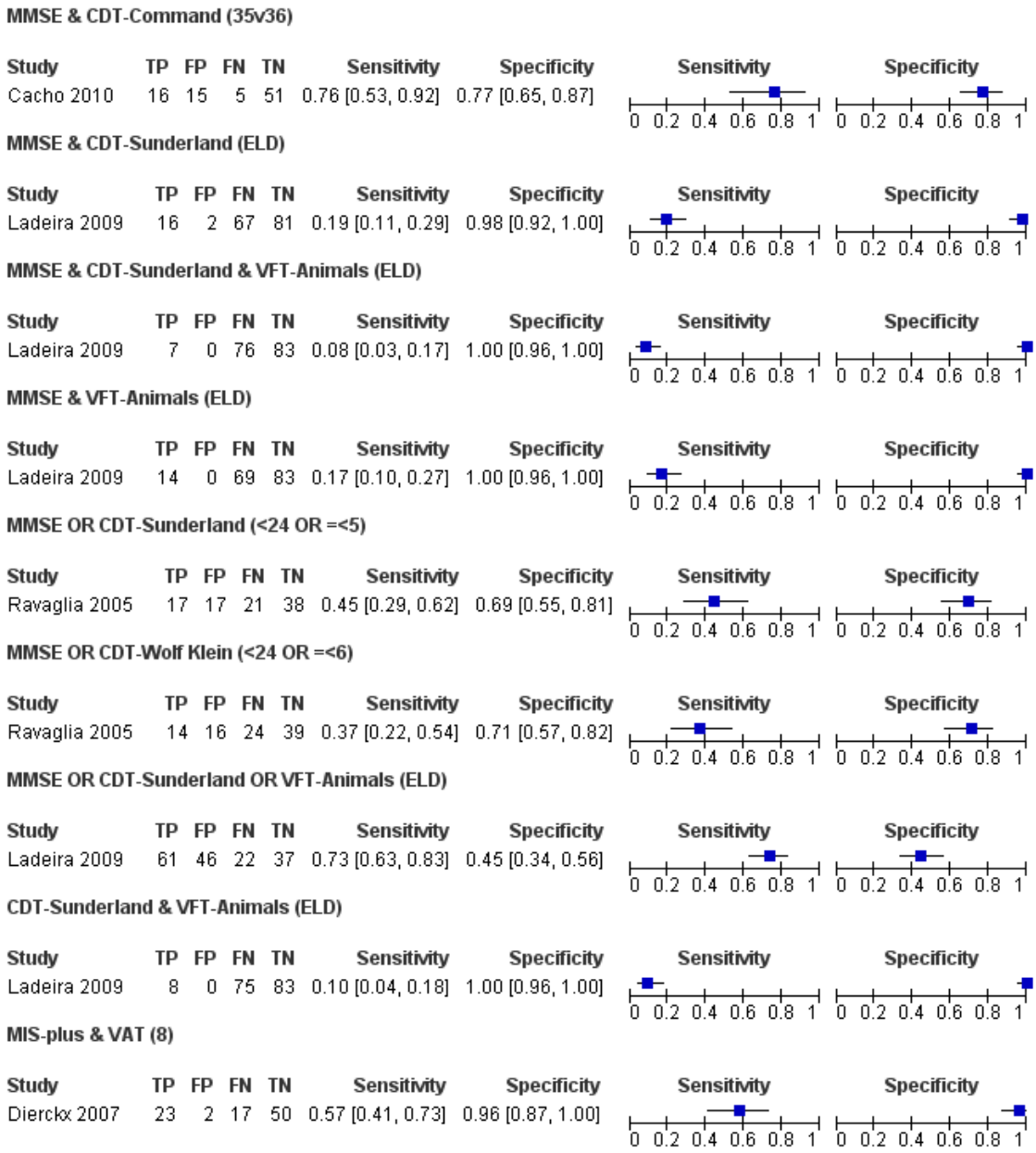


NB: cut-off score in brackets; for index test definitions, see Abbreviations

Appendix 2.5: Forest plot of diagnostic odds ratios of studies looking at the diagnostic accuracy of MoCA (cut off score 23/24)

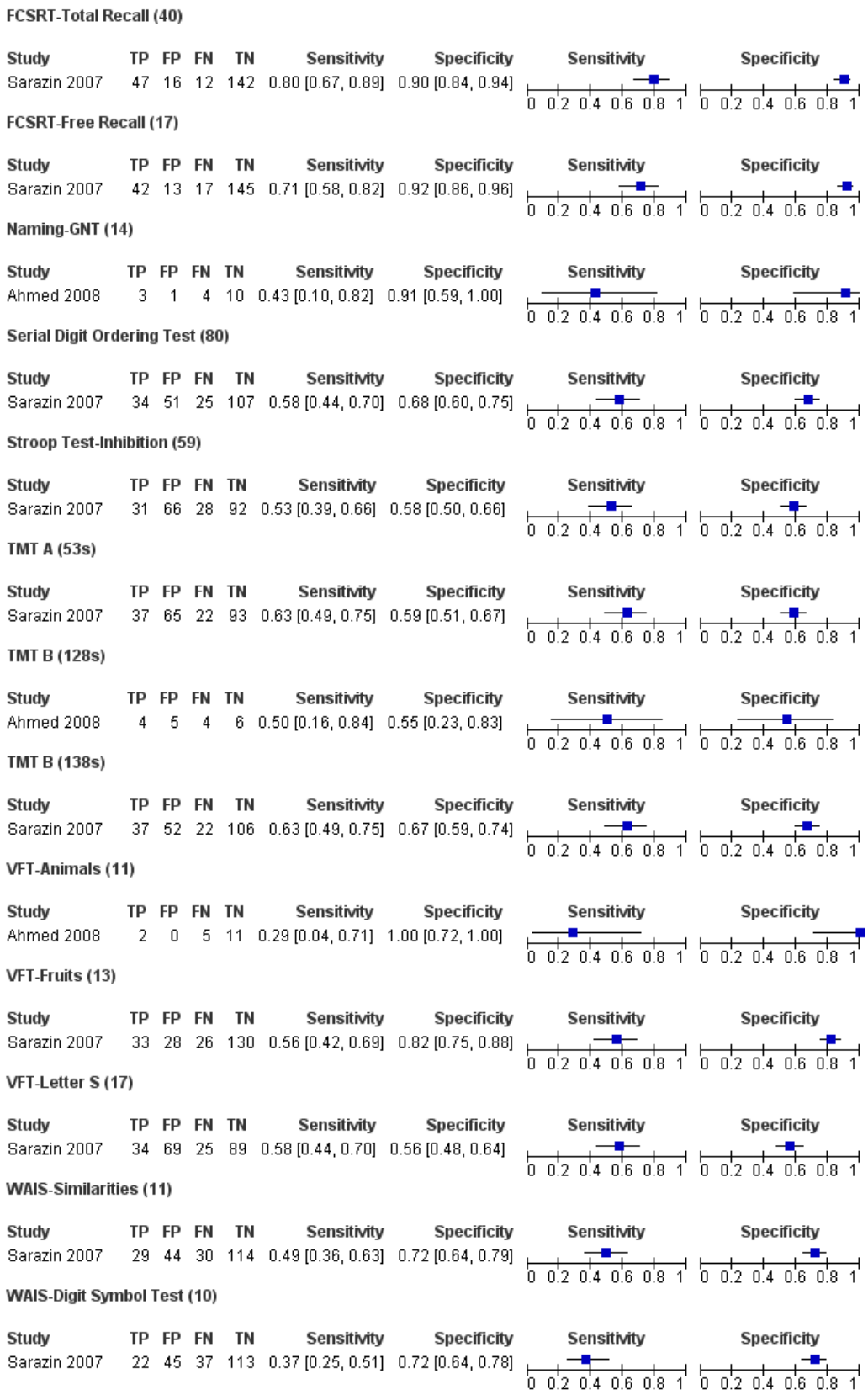


Appendix 2.6: Forest plots summarising sensitivity and specificity for combined screens for identifying aMCI, with associated 95% confidence intervals



NB: cut-off score in brackets; for index test definitions, see Abbreviations

Appendix 2.7: Forest plots summarising sensitivity and specificity for single task screens for identifying aMCI-progressors, with associated 95% confidence intervals



NB: cut-off score in brackets; for index test definitions, see Abbreviations

Appendix 2.8: Risk of Bias

Study	Selection of Patients		Index Test	Ref Standard	Patient Flow			Overall
	1	2	3	4	5	6	7	RoB
(Ahmed et al., 2012)	YES	U/C	U/C	YES	U/C	U/C	YES	U/C
(Ahn et al., 2010)	NO	U/C	NO	NO	U/C	NO	YES	HIGH
(Alegret et al., 2009)	NO	U/C	U/C	YES	U/C	YES	YES	HIGH
(Cacho et al., 2010)	NO	YES	YES	YES	U/C	YES	YES	HIGH
(Chandler et al., 2005)	NO	U/C	U/C	YES	U/C	NO	YES	HIGH
(Costa et al., 2012)	NO	U/C	U/C	U/C	U/C	NO	YES	HIGH
(Dierckx et al., 2007)	U/C	U/C	U/C	U/C	U/C	NO	YES	HIGH
(Diniz et al., 2008)	NO	U/C	U/C	U/C	YES	YES	YES	HIGH
(Freitas et al., 2013)	NO	U/C	NO	NO	YES	NO	YES	HIGH
(Fujiwara et al., 2010)	NO	U/C	YES	YES	YES	NO	YES	HIGH
(Gonzalez-Palau et al., 2013)	NO	U/C	YES	YES	U/C	U/C	YES	HIGH
(Guo et al., 2012)	NO	U/C	U/C	U/C	U/C	U/C	YES	HIGH
(Hanyu et al., 2009)	U/C	U/C	U/C	U/C	U/C	NO	YES	HIGH
(Hanyu et al., 2011)	NO	U/C	U/C	U/C	U/C	YES	YES	HIGH
(Karrasch et al., 2005)	NO	U/C	U/C	U/C	U/C	YES	NO	HIGH
(Kato et al., 2013)	NO	U/C	U/C	U/C	U/C	NO	YES	HIGH
(Ladeira et al., 2009)	YES	U/C	U/C	YES	U/C	YES	YES	U/C
(Lee et al., 2008)	YES	U/C	YES	YES	U/C	U/C	YES	U/C
(Loewenstein et al., 2009)	NO	U/C	U/C	U/C	U/C	U/C	YES	HIGH
(Luis et al., 2009)	YES	U/C	YES	YES	U/C	U/C	YES	U/C
(McLennan et al., 2011)	YES	YES	YES	YES	YES	YES	YES	LOW

Study	Selection of Patients		Index Test	Ref Standard	Patient Flow			Overall RoB
	1	2	3	4	5	6	7	
(Muangpaisan et al., 2010)	U/C	U/C	U/C	U/C	U/C	U/C	YES	U/C
(Nasreddine et al., 2005)	NO	U/C	YES	YES	U/C	YES	YES	HIGH
(Rahman & El Gaafary, 2009)	YES	U/C	YES	YES	U/C	YES	YES	U/C
(Rami et al., 2007)	NO	U/C	U/C	U/C	U/C	NO	YES	HIGH
(Rami et al., 2010)	YES	U/C	U/C	U/C	U/C	U/C	YES	U/C
(Ravaglia et al., 2005)	YES	U/C	NO	NO	YES	YES	YES	HIGH
(Saka et al., 2006)	NO	U/C	U/C	YES	YES	NO	YES	HIGH
(Scheurich et al., 2005)	YES	YES	U/C	U/C	U/C	YES	YES	U/C
(Schrijnemaekers et al., 2006)	YES	U/C	U/C	YES	NO	YES	YES	HIGH
(Smith et al., 2007)	YES	U/C	YES	U/C	U/C	YES	YES	U/C
(Takahashi et al., 2012)	YES	U/C	YES	YES	U/C	YES	U/C	U/C
(Tsai et al., 2012)	NO	U/C	YES	U/C	YES	U/C	YES	HIGH
(Woodard et al., 2005)	YES	U/C	U/C	U/C	YES	YES	YES	U/C
(Yoshida et al., 2012)	NO	U/C	U/C	U/C	U/C	U/C	YES	HIGH
(Zhao et al., 2011)	NO	U/C	U/C	YES	YES	YES	YES	HIGH
(Zhao et al., 2012)	NO	U/C	U/C	U/C	U/C	YES	YES	HIGH
Longitudinal Studies								
(Ahmed, Mitchell, Arnold, Nestor, et al., 2008)	NO	YES	NO	NO	N/A	YES	NO	HIGH
(Sarazin et al., 2007)	YES	U/C	YES	U/C	N/A	YES	NO	HIGH

KEY: 1 = Was a case-control design avoided?; 2 = Was a consecutive or random sample of patients enrolled?; 3 = Were the index test results interpreted without knowledge of the results of the reference standard?; 4 = Were the reference standard results interpreted without knowledge of the results of the index test?; 5 = Was there an appropriate interval between the index test and reference standard?; 6 = Did all patients receive the same reference standard?; 7 = Were all patients included in the analysis? ; RoB = Risk of Bias; U/C = unclear; N/A = not applicable

Appendix 3.1: Recruitment Flyer (v2)

Bradford Teaching Hospitals NHS
NHS Foundation Trust

Are you more forgetful than you used to be?

We need your help with our memory research

You have been contacted because you are aged 70 years or over.

We are testing two new short questionnaires to see if they are as good as a longer assessment currently used when trying to identify people with memory problems.

These shorter questionnaires may be useful in research and may also help GPs in the future to identify people with memory problems.

Please turn over to find out how you can help...

*This research is being managed by **Professor John Young** and **Miss Seline Ozer**, based at the Bradford Institute for Health Research, Bradford Royal Infirmary*

Participant ID: MCI Study_Recruitment Flier_v0.9_221113

Date received:

Taking part will involve a researcher visiting you at home, followed by a visit to our research offices at Bradford Royal Infirmary *(we can help you with travel if required, and any travel expenses will be reimbursed)*

If you are interested in taking part, please answer the following questions (***circle yes or no***) and return the completed leaflet to us using the stamped addressed envelope enclosed:

Please note: your answers will only be seen by the research staff.

1. Do you have any difficulty with your memory?	**YES**	**NO**
2. Do you speak English?	**YES**	**NO**
3. Did you attend school for at least 8 years?	**YES**	**NO**
4. Do you have a husband/wife/relative/friend who would be willing to answer some questions during the study?	**YES**	**NO**

(Return to: Miss S Ozer, Bradford Institute for Health Research, Temple Bank House, Bradford Royal Infirmary, Duckworth Lane, Bradford, BD9 6RJ)

Thank you for taking the time to read and complete this leaflet—we appreciate your help. Please write your details below so that we can contact you further about this study:

Participant ID:

NAME:	
TEL. NO:	
ADDRESS:	

Date received: GP ID:

Appendix 4.1: Memory Alteration Test (M@T)

Encoding

'Try to remember these words. It is important to pay close attention'

Repeat please: **cherry (R) axe (R) elephant (R) piano (R) green (R)**

1	I told you the name of a fruit, what was it?	0	1
2	I told you the name of a tool, what was it?	0	1
3	I told you the name of an animal, what was it?	0	1
4	I told you the name of a musical instrument, what was it?	0	1
5	I told you the name of a colour, what was it?	0	1

(for each question, if 0, repeat the correct answer)

'Later on I will ask you to recall these words'

'Please pay attention to these sentences and try to remember them' (max 2 trials):

Please repeat: **Thirty grey cats ate all the cheese (R)**

6	How many cats were there?	0	1
7	What colour were they?	0	1
8	What did they eat?	0	1

(If 0 tell the subject the correct answer)

Please repeat: A boy named Louis was playing with his bicycle (R) *(max 2 trials):*

9	What was the boy's name?	0	1
10	What was he playing with?	0	1

(If 0 tell the subject the correct answer)

Encoding Score ____

Temporal orientation

11	Day of the week	0	1
12	Month	0	1
13	Date	0	1
14	Year	0	1
15	Season	0	1

Orientation Score ____

Semantic memory

(2 trials; if the subject is wrong, repeat the question)

16	What is your date of birth?	0	1
17	What do you call someone who repairs cars?	0	1
18	What was the name of the last prime minister*?	0	1
19	What is the last day of the calendar year*?	0	1
20	How many days are there in a year?	0	1
21	How many grams are there in one quarter of a kilo?	0	1
22	What is the 8th month of the year?	0	1
23	When is Christmas day?	0	1
24	If the clock shows 11 o'clock, what number does the long hand point toward?	0	1
25	Which season comes after summer?	0	1

Continued overleaf…

26 In the story of Adam and Eve, which animal deceived Eve with an apple*? 0 1

27 Which fruit is necessary to make wine? 0 1

28 Which plant is necessary to make chocolate? 0 1

30 How many hours are there in two days? 0 1

Semantic Memory Score

Free recall

31 Tell me all the words you can remember from the list I said to you at the start of the test

(1 point per word: cherry-axe-elephant-piano-green)

(Wait for the answer minimum 20 sec. You may repeat the question twice)

32 Do you remember anything from the sentence about the cats?

(1 point per idea:30–Gray–cheese)

33 Do you remember anything from the sentence about a boy?

(1 point per idea: Louis–cycle)

Free Recall Score

Cued-recall (if applicable)

(Score 1 point for each word provided in the preceding question)

34 I told you the name of a fruit, what was it? 0 1

35 I told you the name of a tool, what was it? 0 1

36	I told you the name of an animal, what was it?	0	1
37	I told you the name of a musical instrument, what was it?	0	1
38	I told you the name of a colour, what was it?	0	1

Try to remember the sentence about cats . . .

39	How many cats were there?	0	1
40	What colour were they?	0	1
41	What did they eat?	0	1

Try to remember the sentence about a boy . . .

42	What was the boy's name?	0	1
43	What was he playing with?	0	1

Cued Recall Score ____

GLOBAL Score ____

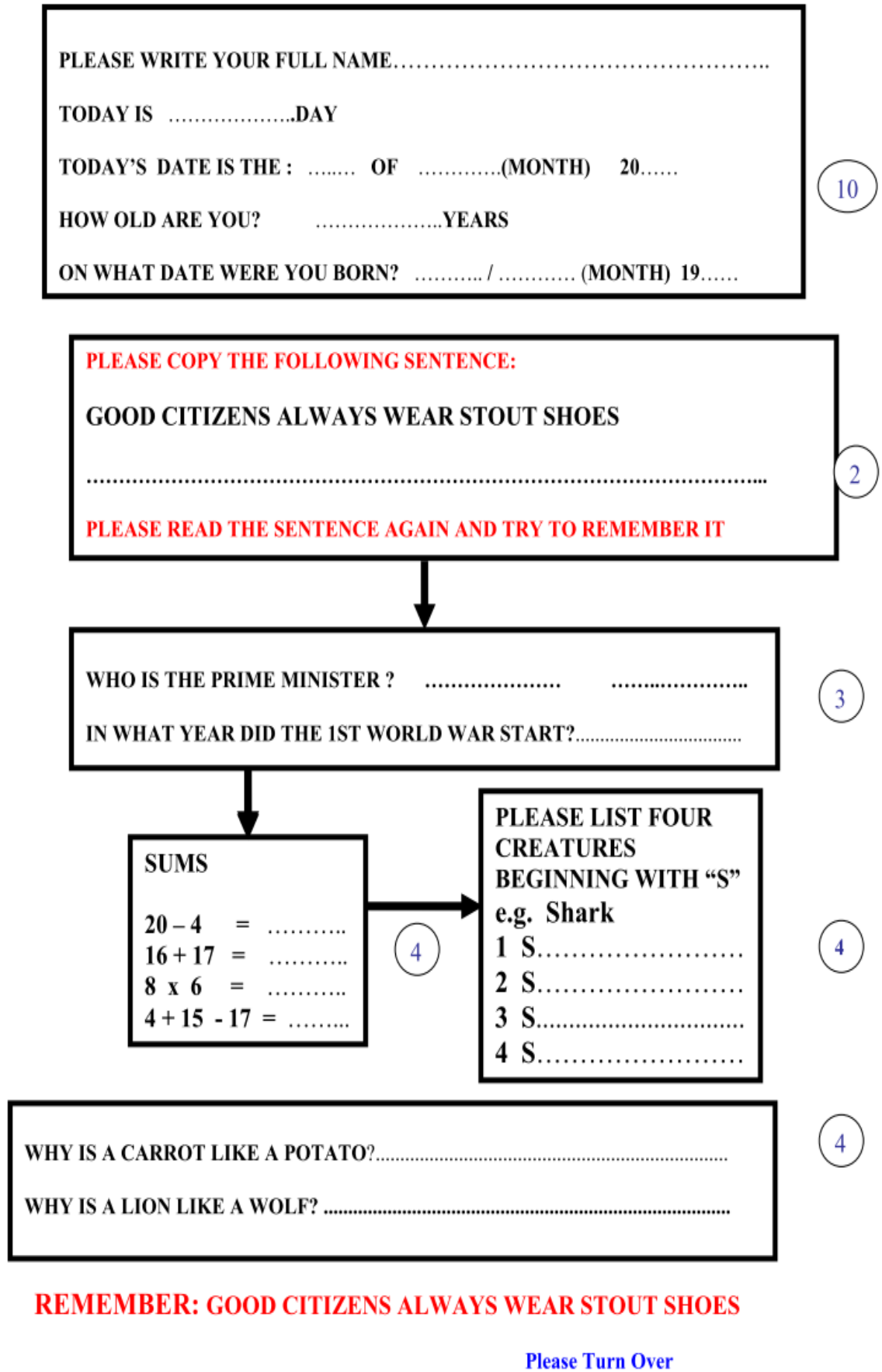
Time to complete ____

** Non-validated English version*

Appendix 4.2: Changes to M@T questions

Original Question	Changed to
What was the name of the last president?	What was the name of the last prime minister?
What is the last day of the year?	What is the last day of the calendar year?
In the Bible, which animal deceived Eve with an apple?	In the story of Adam and Eve, which animal deceived Eve with an apple?
What is the triple of one?	What is three times one?
How many words that I said at the beginning can you remember?	Tell me all the words you can remember from the list I said to you at the start of the test

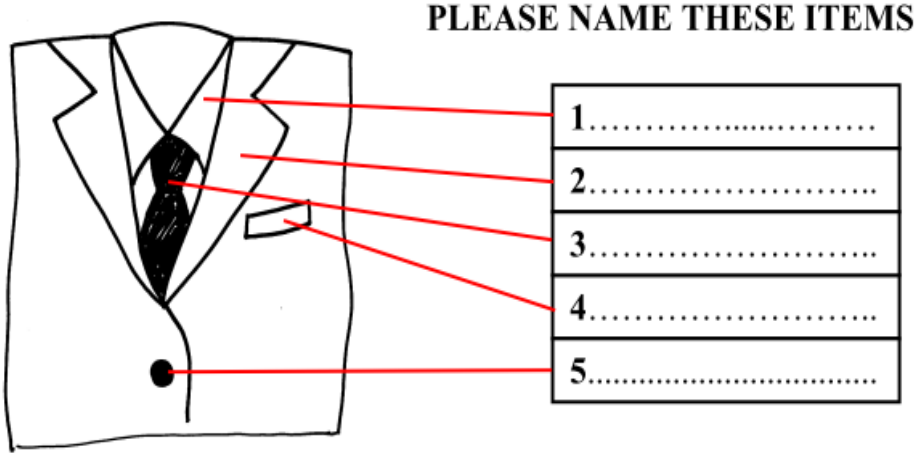
Appendix 4.3: Test Your Memory test (TYM)



REMEMBER: GOOD CITIZENS ALWAYS WEAR STOUT SHOES

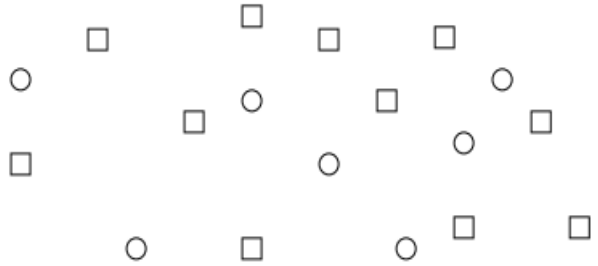
Please Turn Over

PLEASE NAME THESE ITEMS



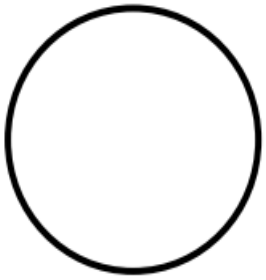
5

PLEASE JOIN THE CIRCLES TOGETHER TO FORM A LETTER (IGNORE THE SQUARES)



3

PLEASE DRAW IN A CLOCK FACE, PUT IN THE NUMBERS 1 – 12 AND PLACE THE HANDS AT 9.20



4

WITHOUT TURNING BACK THE PAGE, PLEASE WRITE DOWN THE SENTENCE YOU COPIED EARLIER :

.....

6

.....

FOR THE TYM TESTER:
HELP GIVEN: NONE/TRIVIAL/MINOR/MODERATE/MAJOR

5

TICK BOX IFANSWERS WRITTEN FOR PATIENT ☐

/50

Appendix 4.4: Demographic characteristics and M@T & TYM scores of matched aMCI and Control groups

	aMCI (n=40)			Control (n=112)			U	z	p value
	Mean (SD)	Median (IQR)	Mean Rank	Mean (SD)	Median (IQR)	Mean Rank			
Age, years	78.0 (5.4)	78 (9)	80	77.4 (5.1)	77 (8)	75	2107.0	-0.6	0.58
Education, years	11.4 (2.0)	11 (2)	74	11.5 (1.9)	11 (2)	77	2141.5	-0.4	0.67
NART IQ	111.4 (9.9)	113 (17)	75	111.8 (9.8)	112.5 (15)	77	2175.5	-0.3	0.79
M@T global score	33.9 (6.2)*	35 (9)*	30	43.4 (4.6)†	44 (6)†	91	397	-7.6	<0.001
TYM global score	40.9 (4.7)	41 (6)	43	45.6 (3.2)	46 (4)	89	893	-5.7	<0.001

KEY: IQR = interquartile range; SD = standard deviation; NART = National Adult Reading Test; M@T = Memory Alteration Test; TYM = Test Your Memory test

p values reported are for aMCI vs. Control

*n = 39; †n = 110