**Improving the accuracy of brief cognitive assessments when used as part of the process for identifying dementia in general practice**

Submitted by Harriet Hunt to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Medical Studies

In November 2018



I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..……..

# Thesis abstract

Identifying dementia in general practice remains a considerable challenge, with mild to moderate stages of dementia potentially underdiagnosed in 30-50% of cases. The primary aim of this PhD thesis was to address the question “how can we improve the accuracy of brief cognitive assessments when used as part of the process for identifying dementia in general practice?”. This was carried out via a combination of secondary research through three evidence syntheses, and primary research via a survey of general practitioners with results triangulated with existing research and thesis findings.

Through the conduct of a rapid review of clinical practice guidelines (CPGs),I found a lack of consistent recommendations for general practice regarding selection and application of brief cognitive assessment (BCA) tools There was also a paucity of guidance given within the identified CPGs on tailoring BCA choice and use for specific populations. The rapid review indicates that greater clarity and consistency is needed from CPGs relating specifically to the use of BCAs as part of the process for identifying dementia in general practice. The systematic review and overview identified an absence of existing evidence. Where evidence exists, BCAs performed inconsistently and were broadly inadequate as a tool for use in general practice dementia care. Other factors beyond diagnostic accuracy render established tests ill-suited for general practice such as administration time, cost and acceptability for clinicians and patients. A number of areas are identified both in cognitive testing and research methods where progress can be made relatively simply.

This thesis demonstrates that many assumptions underlying current practice are without robust foundations, with severe implications for general practice and patient care at a time of scarce resource and growing demand. These assumptions need revising as a priority. What is needed is clear, specific, well-designed primary research to begin to unpick these complexities and realistically address the challenges presented by the identification of dementia within general practice and primary care. I provide explicit recommendations for the design and conduct of a primary comparative accuracy study alongside a trial of effectiveness and cost-effectiveness of using brief cognitive assessments as part of the process of identifying dementia in general practice in order to objectively and systematically assess the suitability of brief cognitive assessments as a tool for use in this population and setting.

**Table of Contents**

Thesis abstract ..... 3
Glossary ..... 9
1. Background ..... 11
1.1. Background to the thesis question ..... 11
1.2. Dementia: its natural history and management ..... 12
1.3. Cognitive assessment in the assessment of dementia ..... 14
1.4. Evaluation of diagnostic tests ..... 17
Glossary ..... 25
2. Aims and objectives ..... 27
2.1. Aim and objectives of this PhD research ..... 27
2.2. Background to the thesis question ..... 27
2.3. Thesis structure ..... 29
2.4. Publications and scholarship activities arising from this thesis ..... 30
Glossary ..... 33
3. A rapid review of clinical practice guidelines relating to the use of brief cognitive assessments used as part of the process for diagnosing dementia within general practice ..... 35
3.1. Why do a rapid review of clinical practice guidelines? ..... 35
3.2. What do we mean by 'brief cognitive assessment'? ..... 37
3.3. Methods ..... 43
3.4. Results ..... 45
3.5. Discussion/ Conclusion ..... 52
Glossary ..... 57
4. An overview of systematic reviews summarising the accuracy of brief cognitive assessments for identifying dementia in primary care ..... 60
4.1. What is an overview? ..... 60
4.2. Method ..... 63
4.3. Search methods for identification of reviews ..... 65
4.4. Overall results ..... 66
4.5. Discussion and relevance ..... 90
4.6. Summary of suggested improvements ..... 100
4.7. Overview review team acknowledgements ..... 103
Appendices ..... 111
Glossary ..... 134
5. How accurate are GPCOG and MMSE in identifying dementia when directly compared to each other? ..... 136
5.1. Background to the review ..... 136

5.2. Methods ..... 139
5.3. Results ..... 148
5.4. Main findings ..... 178
5.5. Discussion ..... 181
5.6. Recommendations ..... 190
Glossary ..... 207
6. The clinical reality of identifying dementia using brief cognitive assessments as part of the primary care consultation ..... 210
6.1. Background: why this survey was needed ..... 210
6.2. Methods ..... 213
6.3. Survey analysis ..... 215
6.4. Free text responses ..... 216
6.5. Overview of responses ..... 216
6.6. Brief cognitive assessment selection and use ..... 219
6.7. Brief cognitive assessment use as part of GP decision making to identify dementia ..... 220
6.8. Confidence in practice ..... 224
6.9. Analysis of free test responses within the questionnaire ..... 226
6.10. Barriers and facilitators for practice identified by respondents ..... 227
6.11. Summary of main findings ..... 232
6.12. Threats to validity ..... 233
6.13. Triangulation of the data, incorporating perspectives from other parts of the thesis and existing literature ..... 234
6.14. Recommendations to practice ..... 239
6.15. Recommendations to research ..... 240
7. Discussion ..... 282
7.1. Aim and objectives ..... 282
7.2. Chapter summaries ..... 284
7.3. Comparison with what is currently known ..... 287
7.4. Limitations ..... 294
7.5. What does this PhD thesis add? ..... 311
7.6. Reflections on the PhD ..... 316
7.7. Thanks and acknowledgements ..... 317
References ..... 321

**Table of Figures**

Figure 1. Standard diagnostic pathway ..... 13
Figure 2. PRISMA flow diagram for rapid review of CPGs ..... 46
Figure 3 PRISMA flow diagram of overview ..... 74
Figure 4 Flow chart of studies identified via the overview ..... 142
Figure 5 Flow chart of all identified studies ..... 144
Figure 6. QUADAS-2 Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies ..... 157
Figure 7. QUADAS-2 Risk of bias and applicability concerns: review authors' judgements about each domain for each included study ..... 158
Figure 8. Forest plot displaying sensitivity and specificity of all brief cognitive assessments across studies ..... 164
Figure 9 Paired accuracy of GPCOG Total and MMSE SROC plot from studies that reported MMSE at the <24 threshold [Basic 2009, Brodaty 2016 and Li 2013] ..... 167
Figure 10 Paired accuracy of GPCOG Total and MMSE SROC plot from studies that reported MMSE at the <25 threshold [Brodaty 2002 and Pirani 2010] ..... 170
Figure 11. Basic 2009 ROC curve comparing RUDAS, MMSE, GPCOG Patient, Informant & Total ..... 171
Figure 12 Brodaty 2002 ROC curve comparing AMT, MMSE, & GPCOG Patient, Informant and Total ..... 172
Figure 13. Pirani 2010 ROC curve comparing MMSE, GPCOG-It Patient, Informant & Total ..... 174
Figure 14. Li 2013 ROC curve comparing Hasegawa's Dementia scale, MMSE, and GPCOG Patient, Informant, Two Step and Total ..... 175
Figure 15. Brodaty 2016 ROC curve comparing MMSE & GPCOG Total ..... 175
Figure 16 Some purposes of medical tests at different stages of the clinical pathway ..... 182
Figure 17 Age of respondents ..... 217
Figure 18 Years since first GP appointment ..... 217
Figure 19 Clustered locations of survey participants by general practice ..... 218
Figure 20 Number of patients assessed for cognitive impairment per month ..... 219
Figure 21 Factors affecting GPs' choice of brief cognitive assessment ..... 220
Figure 22. Have you ever used this brief cognitive assessment for initial assessment? ..... 221
Figure 23. Have you ever used this brief cognitive assessment for monitoring decline over time? ..... 222
Figure 24 Acceptable brief cognitive assessment administration time ..... 223

## Glossary

| | |
|---|---|
| **6-CIT** | 6 Item cognitive impairment test |
| **7MS** | 7 minute screen |
| **AD(D)** | Alzheimer's disease (dementia) |
| **AMSTAR** | Assessment of Multiple Systematic Reviews (quality assessment tool) |
| **AMTS** | Abbreviated Mental Test Score |
| **BIMT/Blessed** | Blessed information Memory Concentration Test |
| **CAMDEX** | Cambridge Mental Disorders of the Elderly Examination |
| **CASI** | Cognitive Abilities Screening Instrument |
| **CDR** | Clinical Dementia Rating |
| **CDT** | clock drawing test |
| **CI** | Confidence interval |
| **CVD** | Cardiovascular disease |
| **DLB** | Dementia with Lewy bodies |
| **DSM** | Diagnostic and Statistical Manual III/III-R/IV/IV-R |
| **DTA** | Diagnostic test accuracy |
| **EMBASE** | Excerpta Medica dataBASE |
| **FN** | False negative |
| **FTD** | Frontotemporal dementia |
| **G8** | Group of Eight (governmental political forum of leading industrial countries) |
| **GMS-AGECAT** | Geriatric Mental State – Automated Geriatric Examination for Computer Assisted Taxonomy |
| **GP** | General Practitioner |
| GPCOG | The General Practitioner assessment of Cognition |
| **ICD** | International Classification of Diseases |
| **IPA-WHO** | International Psychogeriatric Association World Health Organisation criteria |
| **IQCODE** | Informant Questionnaire on Cognitive Decline in the Elderly |
| **m** | Minutes |
| **MCI** | Mild cognitive impairment |
| **MEDLINE** | Medical Literature Analysis and Retrieval System Online |

| | |
|---|---|
| **MMSE** | Mini mental state examination |
| **MSQ** | the Mental Status Questionnaire |
| **NICE** | National Institute for Health and Care Excellence (UK) |
| **NINCDS CERAD** | National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association Consortium to Establish a Registry for Alzheimer's Disease |
| **NINDS-ADRDA** | National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association |
| **NINDS-AIREN** | National Institute of Neurological Disorders and Stroke and Association Internationales pour la Recherche et l´ Ensignement en Neurosciences |
| **NPV** | Negative Predictive Value |
| **NR** | Not reported |
| **NSC** | National Screening Committee |
| **PCL** | Prueba cognitive de leganes *[Leganés cognitive test]* |
| **PPV** | Positive Predictive Value |
| **PRISMA** | Preferred Reporting Items for Systematic Reviews and Meta-Analyse |
| **PROSPERO** | International Prospective Register of Systematic Reviews |
| **PsychInfo** | Database of abstracts of literature in the field of psychology. |
| **QUADAS-2** | Quality Assessment in Diagnostic Accuracy Studies |
| **ROBIS** | Risk of Bias in Systematic Reviews |
| **RUDAS** | The Rowland Universal Dementia Assessment Scale |
| **s** | Seconds |
| **SASSI** | Short And Sweet Screening Instrument |
| **SPMSQ** | Short portable mental status questionnaire |
| **TN** | True negative |
| **VaD** | Vascular Dementia |

# 1. Background

*There is no such thing as a single accurate test at present. Currently GPs would use a process of exclusion and series of cognitive tests over time to reach a diagnosis*

- ARUK submitted evidence to the All-Party Parliamentary Group on Dementia 2012 Unlocking Dementia report

This chapter introduces the background to the thesis question "how can we improve the accuracy of brief cognitive assessments when used as part of the process for identifying dementia in general practice?" and indicates why this question is important.

After describing the syndrome of dementia, its natural history and management with particular focus on diagnosis, cognitive assessment and dementia diagnosis in primary care is discussed, alongside the main tools for diagnosis and evaluation of diagnostic tests.

## 1.1. Background to the thesis question

Dementia is a substantial public health problem which will grow with the ageing global population. In 2015 there were an estimated 800,000 people with dementia in the UK and 46.8 million people worldwide[1], with the number set to double by 2040[2]. Dementia is a progressive clinical condition, and symptoms - such as memory loss, language impairment, disorientation and changes in personality[3] - can often be managed[4] but there is currently no cure. What was referred to in the 1980s as a 'silent epidemic'[5,6] is now receiving increased attention, and the focus of governments, pressure groups and charities has increasingly turned to the importance of identifying people with dementia in a timely manner and managing dementia well with medication, health care and social care[7,8]. NHS England figures suggest that in 2016 around 33% of people with dementia in the UK were without a diagnosis[9]. Diagnosis rates have improved in recent years, and there are many potential contributory factors: government-driven financial incentives such as the £55 payments to general practitioners which ran from 2014 to 2016 encouraged a rise in referrals to memory clinics[10,11]. Increased coverage across the media and society, and broad improvements in public health generally may all have contributed[8]. Recent survey findings from a sample of 1,409 people across 5 European countries confirms there remain many practice-driven delays to dementia

diagnosis across Europe with significant barriers to earlier diagnosis presented by carers such as the first professional seen not considering anything was wrong (33%), not believing it was worthwhile pursuing a diagnosis (6.6%) and the person with dementia refusing to seek help (37.9%)[12].

One key stage in the process for addressing dementia is often overlooked, which is the accurate and careful diagnosis of the condition. We do not know how many people are incorrectly identified as having dementia when they do not, or how many people are told they do not have dementia when they do. Figures based on 6% disease prevalence within an average general practice population suggest false positive dementia diagnosis rates (i.e. people told they have dementia when they do not) may be around 10%[13]. Dementia diagnosis is often difficult, particularly in less clearly-differentiated early stages of the syndrome. Patient, family or clinician reluctance to identify a stigmatised condition, perceived lack of treatment options, the belief that problems were part of the normal ageing process, and masking or lack of differentiation with other conditions combine to make dementia diagnosis one of the most challenging and under-recognised complexities of the condition[10,14-22]. In 2013 the James Lind Alliance reported findings from a Priority Setting Partnership carried out with members of the public, clinicians and specialists to identify priorities for dementia research [23]. Number three on the top ten priority list was the question "what is the impact of an early diagnosis of dementia and how can primary care support a more effective route to diagnosis?". Understanding how to improve the use, and specifically the accuracy, of available assessment tools as part of the process for identifying dementia in general practice is central to improving dementia diagnosis. This key challenge is the focus of this thesis. What follows is a closer exploration of what is currently known; the challenges yet to be addressed; and finally an overview of the chapters within this thesis, how they link together and how conclusions are drawn.

## 1.2. Dementia: its natural history and management

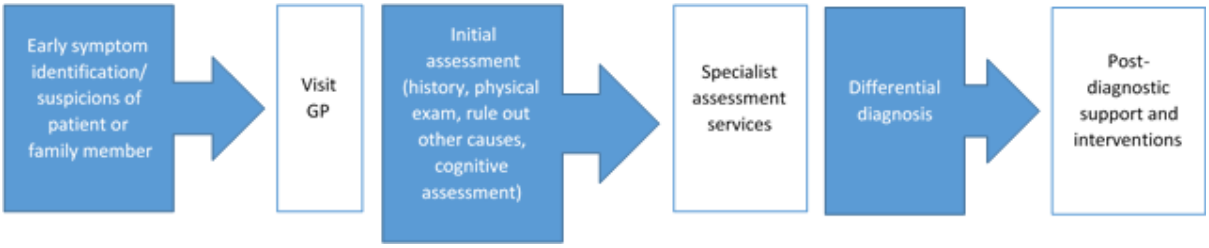
'Dementia' is a broad umbrella term used to describe a set of symptoms including memory decline, other cognitive deficits such as the inability to make judgements, plan ahead and organise, and other changes in social behaviour such as emotional instability, irritability, apathy or disinhibition. A number of other conditions can cause similar symptoms of cognitive deficits, such as heavy alcohol consumption,

depression, urinary infections and thyroid problems, so these changes should be sustained for over 6 months to avoid misclassifying another condition as dementia.

There is no single test for dementia. Instead, the pattern of symptoms that a person has, combined with the history of the individual, talking to friends and relatives who know the person, expert clinical judgment and other investigations such as brief cognitive assessments, neuropsychological tests, blood tests and brain scans will help to diagnose an individual with dementia. Differential diagnosis normally takes place in a specialist memory clinic with the aid of geriatricians, neuropsychologists and psychiatric assessment.

In most European models of care, two-step triage and diagnostic evaluation pathway for dementia is preferred, as endorsed by the European Dementia Consensus Network[24]. Here, the primary care practitioner conducts preliminary assessments including cognitive screening tests to identify the presence and extent of any impairment. Once a positive identification is made the patient is referred on to secondary specialist services for comprehensive assessment and diagnosis. Following full diagnosis, the patient is referred back to the primary care practitioner for ongoing care[25] (see **Error! Reference source not found.** below).

Figure 1. Standard diagnostic pathway




Once dementia has been identified and other conditions such as those mentioned above have been ruled out, there are limited treatment options for the individual, which may help to slow cognitive decline and improve quality of life. Cost-effective drug therapy such as acetylcholinesterase inhibitors may be available, although clinical effectiveness depends on a number of factors including stage of disease progression and dementia sub-type identified. Non-drug interventions such as cognitive stimulation therapy are also available in many areas[26]. Management of the condition is normally through a combination of therapeutic treatment, information provision, signposting and practical support. Carer burden is increasingly recognised as an associated issue in dementia management, as carers are at additional and increased risk of physical and

mental illness as a consequence of the caring role[12]. Social and legal implications also require management for individuals and their carers, such as the ability to drive, loss of independence and legal autonomy[27].

A 2012 survey by the Alzheimer's Society[28] found that 68% of respondents with dementia waited longer than a year between noticing symptoms and getting a diagnosis of dementia, and research has clearly shown that a delay in diagnosis is a clear concern to patients[29,30]. There is some evidence that screening as part of dementia diagnosis can bring an unjustifiable degree of stress to the individual and their families[31], yet other evidence shows that patients and caregivers will already be distressed if progressive cognitive decline is present and it would be naive to think that the diagnostic process itself is a cause of this stress[32]. Indeed, many patients and caregivers report finding relief in a diagnosis[33,34]. A recent survey of 446 Australian outpatients found that 92% of respondents preferred a diagnosis of dementia to be disclosed as soon as possible, with 88% of respondents preferring disclosure as soon as possible in a spouse or partner[35].

The process of diagnosis can identify a dementia syndrome (a set of symptoms such as memory problems, cognitive impairment and changes in social behaviour for at least 6 months), sometimes referred to as 'syndromal diagnosis'[36-38]. Subsequent specific diagnostic work-up ('aetiological diagnosis') can differentiate between dementia subtypes such as Alzheimer's disease, vascular dementia, dementia with Lewy bodies, fronto-temporal dementia and mixed dementia. This more detailed diagnostic investigation is generally carried out by neurological and gerontological experts in e.g. memory clinics following GP referral. The interaction between referral pathways can be problematic and often unclear across many countries[39].

## 1.3. Cognitive assessment in the assessment of dementia

Cognitive assessment concerns investigation of the higher cortical functions, namely memory, attention, orientation, language, planning activities, and sequencing of activities[40]. Cognitive assessment is often positioned as an iterative process comprising various stages, in contrast to a single point valuation. This correctly reflects the clinical reality of cognitive assessment as a process, but is a challenge when considering the purpose of a brief cognitive assessment, and the population in which it may be used. Part of the set of broadly low cost and easily-accessed tools sometimes colloquially referred to as 'pen and paper tests'[41], brief cognitive

assessments measure cognitive function to aid decision making for diagnosis, staging of severity, further treatment, management or onwards referral. The addition of 'brief' is to indicate the tests are relatively quick to administrate, and are generally understood to take up to 10 minutes[42,43]. This description of 'brief' is arbitrary and problematic, and will be discussed fully later in this thesis. Brief cognitive assessments are often used in primary health care such as general practice to provide information on a patient's cognitive function and are part of the recommended toolkit for clinical assessment of dementia within all main national and international guidelines[26,39,40,44-46]. Cognitive assessment is useful in assessing many conditions, including dementia, which is commonly indicated by a decline in cognitive function. As part of the diagnostic process, results from these assessments may guide the general practitioner in their decisions on clinical management, treatment and further testing.

General practice is often the first place people with concerns about their memory will seek help[27]. Across many health care systems (including the USA, Netherlands, the UK, Spain, Italy and Ireland), the referral pathway for dementia diagnosis incorporates general practitioners (GPs) in a 'gatekeeper' role[39,47,48]. The GP or family doctor conducts initial assessments and then refers on as appropriate to specialist services for further diagnostic work up and differential diagnosis. Increasingly, policymakers at both local and international levels have tried to influence diagnostic practice - particularly at the entry level of primary care - in an attempt to increase dementia diagnosis rates[10,15,49-52]. As dementia has received increased focus as a growing social and public health issue, governments have sought to increase dementia diagnosis rates within general practice using financial incentives and clinical practice guidelines[2,9,10]. An interesting question is to what extent these guidelines are consistent in their advice, and how do the various guidelines match current evidence.

Guidelines and recommendations around tests such as brief cognitive assessments should ultimately be based on their ability to improve patient outcome. This is difficult in practice where information on the direct effect of tests on patient outcome is rare, and common measures relate to aspects of test performance such as diagnostic accuracy. The diagnostic accuracy of a test (how well it correctly identifies people with and without a disease) relies on many different pieces of information. Influential factors include disease prevalence, a reliable reference standard, clear patient characterisation and an understanding of the setting the test will be used in – as well

as a strong sense of how the test will be used to guide decision making. A test should also improve patient outcome, it is important to consider what other aspects of a test's performance may influence outcome beyond test accuracy, such as how long it takes to administer, the cost of a test and how acceptable it is to patients[53].

The challenge of testing is heightened when the condition being investigated has a mixed presentation, challenging aetiology, and is hard to identify in the early stages as is the case with dementia[14]. As dementia is syndromic (i.e. a collection of symptoms), there are some symptoms which manifest in people early on and some which will never appear in others.

Many people have more than one type of dementia-causing condition[54-56]. Alzheimer's disease is identified in 60-80% of dementia cases, and vascular dementia makes up 10-17% of cases. Other less common subtypes exist such as frontotemporal dementia, dementia with Lewy bodies and Parkinson's dementia and these have varied underlying mechanisms which can drive different symptomatic expressions[57-59]. As these symptoms can vary widely, this presents a challenge for brief cognitive assessment aiming to capture a range of presentations[43].

The inadequacy of current tools was recognised in a 2012 report "Unlocking dementia" by the UK All-Party Parliamentary Group on Dementia[60]. In submitted evidence, Alzheimer's Research UK stated:

There is no such thing as a single accurate test at present. Currently GPs would use a process of exclusion and series of cognitive tests over time to reach a diagnosis[60].

Poor performance of assessment tools was acknowledged as a potential barrier to dementia diagnosis, and government ministers recommended that "GP training on dementia diagnosis should contain information on known problems with assessment tools and encourage GPs to use their clinical judgement". Moreover, the All Party Parliamentary Group noted that "GPs should feel confident to use their clinical judgement in addition to assessment tools, particularly where there are known problems with the tools"[60].

There have been few developments addressing the performance of brief cognitive assessment tools, yet there is continuous guidance issued by charities, pressure groups, medical associations and government organisations on the availability, selection and use of brief cognitive assessments specifically for use in general or

family practice[61-67]. These guidelines, how they harmonise and diverge and the evidence base upon which they are established are discussed within Chapter 3.

Whilst it is undeniably a positive development to recognise the importance of dementia identification, current discourse and strategies appear to have bypassed a central aspect of diagnosis – that is, the performance of the tests being used to assess cognitive function as part of a broader assessment for possible dementia. An overview review of the diagnostic accuracy of brief cognitive assessments used as part of the process for identifying dementia in primary care identifies the current evidence base and is presented in Chapter 4 of this thesis.

Identifying dementia in general practice is a considerable challenge, with mild to moderate stages of the condition potentially underdiagnosed in 30-50% of cases[9,26,29,68-73]. Part of the difficulty encountered in dementia diagnosis – particularly before the condition is advanced - is the challenge in differentiating between dementia subtypes, availability of reliably predictive biomarkers, technology available for use in general practice and recognition of key signs and symptoms that may be masked by other conditions often presenting concurrently as people age. Many initiatives (including the national dementia plans of Australia, Greece, Indonesia, Ireland, Israel, Italy, Luxembourg, Malta, Mexico, Northern Ireland, Scotland, Switzerland, Taiwan, and the USA) focus on improving early or timely identification of dementia to allow space for planning, help-seeking, symptom modification and allocation of resources[26].

As the diagnostic process is moved to an earlier time point in the dementia profile, the challenges presented to clinicians in correctly identifying the signs and symptoms presented are intensified and the likelihood of diagnostic errors (such as false positives, i.e. incorrectly identifying someone as having dementia when they do not) increases. This diagnostic process does not operate in a vacuum, and there are many potential personal, health and social implications for the individual, their friends and families, the clinician and society more broadly which may influence decisions around timely diagnosis of dementia.

## 1.4. Evaluation of diagnostic tests

Diagnostic accuracy is one way of assessing how a ‘new’ test (the index test) compares to the best available test (the reference standard). Diagnostic accuracy is measured as a test’s ability to correctly identify people with a condition (sensitivity)

and to correctly rule out people without a condition (specificity). Sensitivity and specificity are generated by the same data and are therefore interrelated so that – generally – when one increases, the other lowers. In this sense, there is usually a decision to be made about whether it is clinically more meaningful to have higher sensitivity and lower specificity, or higher specificity and lower sensitivity. Diagnostic accuracy is not a fixed element of a test, and factors such as different patient groups, the disease spectrum, the clinical setting, and interpretations placed upon a test all influence accuracy. It is therefore essential to define these aspects up front within the study question, and ideally place the test within context of previous testing strategies and consequent management decisions influenced by test results[74,75].

Diagnostic accuracy data are generated using a binary 2x2 table, where people within a study are either: correctly identified by both the reference standard and the index test (true positives); correctly ruled out by both the reference standard and the index test (true negatives); correctly identified by the reference standard but incorrectly ruled out by the index test (false negatives); and correctly ruled out by the reference standard but incorrectly identified by the index test (false positives). This is illustrated in Table 1 below:

Table 1. Binary 2x2 table classifying diagnostic accuracy

| | | **Reference Standard** | |
|---|---|---|---|
| | | + | - |
| **Index test** | + | True Positive | False Positive |
| | - | False Negative | True Negative |

Whilst much discussion revolves around dementia diagnosis and the need to identify dementia earlier in the diagnostic pathway, there appears to be relatively little discussion of the accuracy of the diagnosis. In their article on the dangers behind new diagnostic tests, Hofmann and Welch highlight problems of tests being applied in different ways, with different populations and earlier in the investigatory process – all issues equally relevant to established tests such as brief cognitive assessments[76]. In the case of dementia, brief cognitive assessments are used as part of the process for measuring cognitive function, and this in turn can indicate the possibility of dementia alongside other causes. Brief cognitive assessments form part of a clinician's toolkit for assessing a patient, and the results of a brief cognitive test should never be viewed in isolation but as part of the broader clinical assessment.

The evaluation of tests in secondary research (such as systematic reviews and overviews) presents particular challenges due to the greater level of abstraction of the test data. Details such as test thresholds, population, sample and disease prevalence and nuances of test presentation can be lost during the construction of this 'second level' evaluation framework. Reporting of test data within secondary research varies considerably, and the candidate has been fortunate to have the opportunity to contribute to improvements in guidelines for the transparent reporting of systematic reviews within the PRISMA-DTA Working Group[77] in order to make the results from systematic reviews of diagnostic test accuracy studies more useful. This work is discussed in more detail in Chapters 2 and 7 in this thesis.

A range of different language is used to describe dementia assessment and diagnosis, and one term commonly used within the literature is 'screening'. This has several meanings depending on who it is applied to, and who is using the term. These issues are addressed in detail below.

Demand for 'timely' diagnosis – that is, diagnosis at the point when cognitive and other changes an individual is experiencing starts to affect their life and lives of people close to them[78] - is strong and improved identification in general practice is generally supported, but there is robust resistance for dementia screening. Although this term is used in different ways, in this thesis screening is used to describe the proactive and asymptomatic investigation for a condition at the level of a population. Whilst screening has been suggested as a way to identify more cases earlier[79], there is little appetite for dementia screening globally for a number of reasons: the low disease prevalence of around 6% within a general practice population; poor assessment tools; and lack of systemic post-diagnostic support[2,60,80,81]. In 2015 the UK National Screening Committee issued clear guidance following a thorough review of the evidence:

> *The UK NSC does not recommend a systematic population screening programme for Dementia using cognitive assessment tools. Cognitive assessment tools are insufficiently reliable for use in large scale screening programmes and there is insufficient evidence on the benefit of interventions to demonstrate that early treatment leads to improved outcomes from screening[82].*

Instead of screening, a case finding approach is preferred by many[13,43,63,68,82-85]. In a case finding approach, people present to their GP with a symptom or set of symptoms, or in 'proactive' case finding individuals who fall into higher risk groups (such as those

over 65 years old, people with pre-existing vascular conditions or Parkinson's disease) are systematically invited for assessment.

When used for case finding, brief cognitive assessments have been shown to identify a higher proportion of people with a positive test result who actually have dementia compared to when they are used for screening[81]. In either case, evidence for the diagnostic accuracy of brief cognitive assessments when used in a primary care population is still very unclear (i.e. how good a test is at correctly identifying someone with dementia, or correctly ruling out someone without dementia).

### 1.4.1. Population screening

Population screening refers to the comprehensive indiscriminate assessment of a given population in order to identify a condition[85,86]. Despite advocacy for screening by a number of test developers[79], major organisations such as the US Preventive Services Task Force[87,88], the US Department of Veteran Affairs[89], Alzheimer's Research UK[85] and the UK National Screening Committee (NSC)[90] do not recommend universal screening for dementia for a number of reasons, most prominent of which is that the way the condition develops and the beneficial effect of early treatment are still uncertain[90]. In addition, the NSC cites evidence of low population prevalence in people over 65 years old (around 7 in 100 people would be affected), and the poor sensitivity of available tests.

Some expert analyses[91-93] have highlighted a lack of evidence on the impact of initiatives designed to increase asymptomatic screening of dementia within primary care such as cost-effectiveness and harms to patients, and this conspicuous uncertainty is worth exploring further in future research.

### 1.4.2. Targeted screening and case finding

In the US, initiation of Medicare Annual Wellness Visits[63] and in the UK the NHS introduction of a Direct Enhanced Service (DES) for dementia assessment raised the profile of dementia screening tests and heightened calls for developing the evidence base for formal dementia screening (including investigating diagnostic thresholds, improving understanding of screening outcomes and further exploration of benefits, harms and costs of screening)[84].

'Targeted screening' or 'case finding' are commonly-used terms referring to the identification of 'cases' or people who may be symptomatic through particular

exposure or higher risk groupings (such as those aged over 65 years old, people who have had a stroke or people over 75 years old admitted to hospital). Case finding for dementia is currently part of the UK Government's Dementia Strategy[94], although supporting evidence is mixed and the approach itself is contentious.

Targeted screening can exploit standard general practice methods such as the regular review of people with long term conditions, and it is argued by some that the investigation of people with increased risk of dementia due to age or related medical conditions would seem a logical part of the primary care assessment[95]. In addition, general practice is a comfortable environment where people with comorbidities and long term conditions will be familiar with both the environment and the health care professionals for regular review. It is argued that this is a familiar, cost-effective and non-stigmatising environment where people at increased risk of dementia and their loved ones can be managed holistically[95].

The case for targeted screening, however, is not met by current evidence according to the Wilson and Jungner criteria applied by the UK National Screening Committee in their 2015 guidelines recommending against dementia screening[90]. In order to meet appropriate and ethical thresholds, these criteria provide four conditions to be met relating to the condition, its diagnosis, treatment, and cost-effectiveness. The condition must be considered important, with a clear understanding of aetiology (the underlying set of causes[96]) contributory and risk factors and observable primary prevention strategies having taken place. The importance of dementia is increasingly recognised, yet evidence on other aspects of aetiology and risk factors remain unclear whilst there is still intense discussion around what forms an early stage of dementia, with associated conditions such as mild cognitive impairment not converting consistently to dementia[32,97].

A clear, safe, precise and validated diagnostic test should be available that is acceptable both to the population at large and the test users, with a known course of therapeutic action if a test result is positive[98]. Systematic review evidence has consistently shown that there are a limited number of brief cognitive assessments available that are demonstrably appropriate or validated for use in primary care settings[99-106].

The process for identifying dementia in general practice is iterative and no single test should be used in isolation to give a definitive answer[7,26,40,48,60,64,65,84,107,108]. As part of the standard clinical assessment, GPs take a medical history from the patient, record signs and symptoms of disorder and explore possible reasons for these signs and symptoms including life events, medication and temporary conditions. Depending on the complaint or observed issues, the GP may run tests to rule out causes such as depression, thyroid dysfunction or urinary infections[62,109].

If dementia or cognitive impairment is suspected, there are a number of brief cognitive assessments recommended by various guidelines including the Mini Mental State Examination (MMSE), General Practitioner Assessment of Cognition (GPCOG), the Mini-Cog, and the Memory Impairment Screen [62,64,110-112]. The criteria for recommending different tests varies, and these tests and guidelines are addressed in detail in Chapter 2 and 3.

As with the majority of tests[86], the diagnostic accuracy of brief cognitive assessments used as part of the process to identify cognitive function in possible dementia cases is imperfect and many factors influence test performance. Factors such as disease prevalence (how many cases of a condition are present within a stated population), setting, intended patient group, test administrator, practice effects and interpretation all have an influence on how well a brief cognitive assessment identifies people with and without cognitive impairment or dementia[7,17,29,39,46,70,113,114].

Many brief cognitive assessments were developed and validated in populations other than general practice[44,115,116]. This has implications for the applicability of stated diagnostic accuracy, as well as for other factors which may influence the suitability and performance of a test such as: administration time; economic cost; accessibility; acceptability to the patient and clinician; and resource availability[42,53,117-119].

There is solid systematic review evidence assessing the diagnostic accuracy of different brief cognitive assessments for identifying dementia[41,42,44,99,116,120]. The picture of which tests are most suitable for general practice, however, remains unclear, and guidelines for diagnosis lack consistent direction for health care professionals, policy makers and the public.

This thesis forms a substantial addition to the current body of knowledge by establishing what is currently known about the diagnostic accuracy of current brief

cognitive assessments when used as part of the process for identifying dementia in primary care. This thesis also progresses methods for conducting overviews of systematic reviews of diagnostic accuracy, and comparative systematic reviews of diagnostic accuracy for specific brief cognitive assessments, as well as bringing clarity to understanding of how accurate current brief cognitive assessments are for identifying dementia in primary care, how useful they are and what recommendations can be made for current use - assessed through a combination of the overview, the systematic review of direct test comparisons, and the GP survey.

## Glossary

| | |
|---|---|
| **6-CIT** | 6 Item cognitive impairment test |
| **7MS** | 7 minute screen |
| **AD(D)** | Alzheimer's disease (dementia) |
| **AMSTAR** | Assessment of Multiple Systematic Reviews (quality assessment tool) |
| **AMTS** | Abbreviated Mental Test Score |
| **BIMT/Blessed** | Blessed information Memory Concentration Test |
| **CAMDEX** | Cambridge Mental Disorders of the Elderly Examination |
| **CASI** | Cognitive Abilities Screening Instrument |
| **CDR** | Clinical Dementia Rating |
| **CDT** | clock drawing test |
| **CI** | Confidence interval |
| **CVD** | Cardiovascular disease |
| **DLB** | Dementia with Lewy bodies |
| **DSM** | Diagnostic and Statistical Manual III/III-R/IV/IV-R |
| **DTA** | Diagnostic test accuracy |
| **EMBASE** | Excerpta Medica dataBASE |
| **FN** | False negative |
| **FTD** | Frontotemporal dementia |
| **G8** | Group of Eight (governmental political forum of leading industrial countries) |
| **GMS-AGECAT** | Geriatric Mental State – Automated Geriatric Examination for Computer Assisted Taxonomy |
| **GP** | General Practitioner |
| **GPCOG** | The General Practitioner assessment of Cognition |
| **ICD** | International Classification of Diseases |
| **IPA-WHO** | International Psychogeriatric Association World Health Organisation criteria |
| **IQCODE** | Informant Questionnaire on Cognitive Decline in the Elderly |
| **m** | Minutes |
| **MCI** | Mild cognitive impairment |
| **MEDLINE** | Medical Literature Analysis and Retrieval System Online |

| | |
|---|---|
| **MMSE** | Mini mental state examination |
| **MSQ** | the Mental Status Questionnaire |
| **NICE** | National Institute for Health and Care Excellence (UK) |
| **NINCDS CERAD** | National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association Consortium to Establish a Registry for Alzheimer's Disease |
| **NINDS-ADRDA** | National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association |
| **NINDS-AIREN** | National Institute of Neurological Disorders and Stroke and Association Internationales pour la Recherche et l´ Ensignement en Neurosciences |
| **NPV** | Negative Predictive Value |
| **NR** | Not reported |
| **PCL** | Prueba cognitive de leganes *[Leganés cognitive test]* |
| **PPV** | Positive Predictive Value |
| **PRISMA** | Preferred Reporting Items for Systematic Reviews and Meta-Analyse |
| **PROSPERO** | International Prospective Register of Systematic Reviews |
| **PsychInfo** | Database of abstracts of literature in the field of psychology. |
| **QUADAS-2** | Quality Assessment in Diagnostic Accuracy Studies |
| **ROBIS** | Risk of Bias in Systematic Reviews |
| **RUDAS** | The *Rowland Universal Dementia Assessment Scale* |
| **s** | Seconds |
| **SASSI** | Short And Sweet Screening Instrument |
| **SPMSQ** | Short portable mental status questionnaire |
| **TN** | True negative |
| **VaD** | Vascular Dementia |

# 2. Aims and objectives

*Diagnosis is the gateway for care. No drug or non-drug treatment can be given, and no specific future planning carried out, without individuals receiving a diagnosis*

Knapp M, Comas-Herera A, Somami A & Banerjee S (2007) Dementia: international comparisons. LSE PSSRU

## 2.1. Aim and objectives of this PhD research

The primary aim of this thesis is to address the question "how can we improve the accuracy of brief cognitive assessments when used as part of the process for identifying dementia in general practice?" by establishing a clear picture of the evidence for diagnostic accuracy of brief cognitive assessments for identifying dementia within general practice.

This is achieved through the following objectives:

- establishing current clinical practice guidelines relating to the use of brief cognitive assessments used to identify dementia in primary care;
- assessing the diagnostic accuracy of brief cognitive assessments used to identify dementia in primary care;
- reviewing the diagnostic test accuracy evidence by assessing studies that have directly compared GPCOG and MMSE for identifying dementia in general practice; and
- exploring the views of general practitioners around dementia screening tests.

This assessment of the evidence allows the building of practical recommendations for improving the accuracy of brief cognitive assessments specifically for use in general practice.

## 2.2. Background to the thesis question

Dementia is a substantial public health problem which will grow with the ageing global population. In 2015 there were an estimated 800,000 people with dementia in the UK and 46.8 million people worldwide[1], with the number set to double by 2040[2]. Dementia is a progressive clinical condition, and symptoms - such as memory loss, language

impairment, disorientation and changes in personality[3] - can often be managed[4] but there is currently no cure. What was referred to in the 1980s as a 'silent epidemic'[5,6] is now receiving increased attention, and the focus of governments, pressure groups and charities has increasingly turned to the importance of identifying people with dementia in a timely manner and managing dementia well with medication, health care and social care[7,8]. NHS England figures suggest that in 2016 around 33% of people with dementia in the UK were without a diagnosis[9]. Diagnosis rates have improved in recent years, and there are many potential contributory factors: government-driven financial incentives such as the £55 payments to general practitioners which ran from 2014 to 2016 encouraged a rise in referrals to memory clinics[10,11]. Increased coverage across the media and society, and broad improvements in public health generally may all have contributed[8]. Recent survey findings from a sample of **1,409 people across 5 European countries** confirms there remain many practice-driven delays to dementia diagnosis across Europe with significant barriers to earlier diagnosis presented by carers such as the first professional seen not considering anything was wrong (33%), not believing it was worthwhile pursuing a diagnosis (6.6%) and the person with dementia refusing to seek help (37.9%)[12].

One key stage in the process for dementia treatment is often overlooked, which is the accurate and careful diagnosis of the condition. We do not know how many people are incorrectly identified as having dementia when they do not, or are told they do not have dementia when they do although figures based on 6% disease prevalence within an average general practice population suggest false positive dementia diagnosis rates (i.e. people told they have dementia when they do not) may be around 10%[13]. Dementia diagnosis is often difficult, particularly in less clearly-differentiated early stages of the syndrome. Patient, family or clinician reluctance to identify a stigmatised condition, perceived lack of treatment options, the belief that problems were part of the normal ageing process, and masking or lack of differentiation with other conditions combine to make dementia diagnosis one of the most challenging and under-recognised complexities of the condition[10,14-22]. In 2013 the James Lind Alliance reported findings from a Priority Setting Partnership carried out with members of the public, clinicians and specialists to identify priorities for dementia research [23]. Number three on the top ten priority list was the question "what is the impact of an early

diagnosis of dementia and how can primary care support a more effective route to diagnosis?". Understanding how to improve the use, and specifically the accuracy, of available assessment tools as part of the process for identifying dementia in general practice is central to improving dementia diagnosis. This key challenge is the focus of this thesis.

## 2.3. Thesis structure

A brief summary of each chapter's content follows:

1. **Chapter 1. Background**
   Background to the thesis question; dementia, its natural history and management, with a particular focus on diagnosis; cognitive assessment in the diagnosis of dementia; and the evaluation of diagnostic tests.

2. **Chapter 2. Aims and objectives**
   Aims and objectives of the thesis; thesis structure.

3. **Chapter 3. Rapid review of clinical practice guidelines relating to the use of brief cognitive assessments used as part of the process for diagnosing dementia within general practice.**
   Why a rapid review of guidelines is needed; research protocol; discussion of the findings, including limitations of included studies and recommendations for practice and research.

4. **Chapter 4**. **An overview of systematic reviews summarising the accuracy of brief cognitive assessments for identifying dementia in primary care**.
   Why is an overview needed; key findings; discussion and recommendations for research and practice.

5. **Chapter 5. Systematic review - how accurate are GPCOG and MMSE in identifying dementia when directly compared to each other?**
   Comparative test performance – methods and results; analysis and discussion of findings; recommendations for practice and further research needed.

6. **Chapter 6. The clinical reality of identifying dementia using brief cognitive assessments as part of the primary care consultation: How do GPs influence diagnostic performance?**
   Survey methods; test selection and GP preferences; barriers and facilitators; other issues identified; summary and discussion.

7. **Chapter 7. Summary and discussion.**
   Synopsis of main findings; recommendations for research and practice; thesis discussion in context; reflection on the PhD.

A glossary is included at the beginning of each chapter, and each chapter is detailed in the page footer along with page numbers to help the reader orientate throughout the thesis.

The referencing style uses EndNote "Academic Medicine", or superscript numbers (e.g.[1]). This is because many chapters, in particular the rapid review in Chapter 3, the overview in Chapter 4 and the systematic review in Chapter 5, are "reference heavy" due to numerous study-level and systematic review-level references. In order to make the text readable superscript numbers are used throughout.

## 2.4. Publications and scholarship activities arising from this thesis

Publications, presentations and wider scholarship activities arising from this thesis, including those in press, are listed below in order of activity type and date:

### 2.4.1. Publications

McInnes, MD, Moher, D, Thombs, BD, McGrath, TA, Bossuyt, PM, Clifford, T, ...& **Hunt, HA** (2018). Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*, *319*(4), 388-396. https://jamanetwork.com/journals/jama/fullarticle/2670259

* PRISMA-DTA Working Group: Tammy Clifford, PhD; Jérémie F. Cohen, MD, PhD; Jonathan J. Deeks, PhD; Constantine Gatsonis, PhD; Lotty Hooft, PhD; **Harriet A. Hunt**, MSc; Christopher J. Hyde, PhD; Daniël A. Korevaar, MD, PhD; Mariska M. G. Leeflang, PhD; Petra Macaskill, PhD; Johannes B. Reitsma, MD, PhD; Rachel Rodin, MD, MPH; Anne W. S. Rutjes, PhD; Jean-Paul Salameh, BSc; Adrienne Stevens, MSc; Yemisi Takwoingi, PhD; Marcello Tonelli, MD, SM; Laura Weeks, PhD; Penny Whiting, PhD; Brian H. Willis, MD, PhD.

**Hunt, HA**, Pollock, A., Campbell, P., Estcourt, L., & Brunton, G. (2018). An introduction to overviews of reviews: planning a relevant research question and objective for an overview. *Systematic Reviews* 7(1), 39.

Pollock A, Campbell P, Brunton G, **Hunt HA** and Estcourt L (2017) Selecting and implementing overview methods: implications from five exemplar overviews. *Systematic Reviews* 6(1), p.145 10.1186/s13643-017-0534 https://doi.org/10.1186/s13643-017-0534-3

**Hunt HA**, van Kampen S, Takwoingi Y, Llewellyn DJ, Pearson M and Hyde CJ, 2017. The comparative diagnostic accuracy of the Mini Mental State Examination (MMSE) and the General Practitioner assessment of Cognition (GPCOG) for identifying dementia in primary care: a systematic review protocol. *Diagnostic and Prognostic Research* 1(1), p.14 https://doi.org/10.1186/s41512-017-0014-1

**Hunt HA** & Hyde CJ (2017) An overview of systematic reviews summarising the accuracy of brief cognitive assessments for identifying dementia in primary care. *Diagnostic and Prognostic Research* 1(Suppl 1):P30

### 2.4.2. Presentations & workshops

**Hunt H**, McGrath T, Salameh J, Dehmoobad Sharifabadi A, Frank R, Whiting P (2018) Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for diagnostic test accuracy studies (PRISMA-DTA): workshop for review authors. Monday 17th September, 11:00 – 12:30 Cochrane Colloquium, Edinburgh 2018.

Weeks L, **Hunt HA**, Pieper D, Hartling L (2018) HTAi Annual Meeting 2018 workshop ID #52 "Overviews Of Systematic Reviews: An Emerging Methodology In Health Technology Assessment" The Westin Bayshore, Vancouver, Canada

**Hunt HA** & Hyde CJ (2017) How to compare medical tests: An example of systematic reviews of diagnostic accuracy using within-study comparisons. Global Evidence Summit 13-16 September 2017, Cape Town, South Africa (abstract accepted for long oral presentation)

**Hunt HA** & Hyde CJ (2017) Overviews of diagnostic test accuracy: more research versus better research. Global Evidence Summit 13-16 September 2017, Cape Town, South Africa (abstract accepted for short oral presentation)

**Hunt HA** & Hyde, CJ (2016) Overview (de)generation: a review of reviews on the diagnostic accuracy of brief cognitive assessments for identifying dementia in primary care. Long oral presentation at *24th Cochrane Colloquium*, Seoul, South Korea

**Hunt HA** (2016) How can primary care support a more effective route to dementia diagnosis? Oral presentation at *University of Exeter Medical School Academic Research Event*, Torquay, UK

Pollock A, **Hunt HA**, Campbell P, Estcourt L and Brunton G (2016) Cochrane overviews of reviews: exploring the methods and challenges. An overview of systematic reviews of diagnostic test accuracy (workshop). Workshop presentation at *Cochrane UK & Ireland Symposium 2016: Impact, Innovation and Ingenuity*. Birmingham, UK

**Hunt H** (2015) Constructing an overview of systematic reviews of diagnostic test accuracy. Oral presentation at 23rd Cochrane Colloquium 2015, Vienna, Austria

### 2.4.3. Scholarship activities

- PRISMA-DTA Consensus Working Group invited member *[2016 – ONGOING]*
- GRADE in Overviews guidance development invited contributor *[2017 – ONGOING]*
- Cochrane Diagnostic Test Accuracy Reviews peer reviewer *[2014 – ONGOING]*
- Cochrane Global Ageing Group member *[2016 – ONGOING]*
- Peer reviewer for British Medical Journal, Diagnostic & Prognostic Research, BMC Medical Research Methodology, BMC Systematic Reviews *[2014 – ONGOING]*

## Glossary

| | |
|---|---|
| **3MS** | Modified Mini Mental Exam |
| **6-CIT** | 6 Item cognitive impairment test |
| **7MS** | 7 minute screen |
| **10-CS** | 10 point Cognitive Screener |
| **ACE** | Addenbrooke's Cognitive Evaluation |
| **ACE-R** | Addenbrooke's Cognitive Evaluation - revised |
| **ACE-III** | Addenbrooke's Cognitive Evaluation – version three |
| **ADAS-Cog** | The Alzheimer's Disease Assessment Scale - Cognition |
| **ADI** | Alzheimer's disease International |
| **AMT(S)** | Abbreviated Mental Test Score |
| **BNT** | Behavioural Neurology Assessment |
| **CCCDTD** | Canadian Consensus Conferences on the Diagnosis and Treatment of Dementia |
| **CCG** | Clinical Commissioning Group |
| **CDPC** | Cognitive Decline Partnership Centre |
| **CDT** | clock drawing test |
| **CPG** | Clinical Practice Guideline |
| **EMBASE** | Excerpta Medica dataBASE |
| **FAB** | Frontal Assessment Battery |
| **FAQ** | Functional Activities Questionnaire |
| **GP** | General Practitioner |
| **GPCOG** | The General Practitioner assessment of Cognition |
| **IQCODE** | Informant Questionnaire on Cognitive Decline in the Elderly |
| **KICA-Cog** | Kimberley Indigenous Cognitive Assessment |
| **KICA-Screen** | Kimberley Indigenous Cognitive screening tool |
| **MEDLINE** | Medical Literature Analysis and Retrieval System Online |
| **MeSH** | Medical Subject Headings |
| **MIS** | Memory Impairment Screen |
| **MMSE** | Mini mental state examination |
| **MoCA** | Montreal Cognitive Assessment |
| **NHS** | National Health Service (UK) |
| **NICE** | National Institute for Health and Care Excellence (UK) |

| | |
|---|---|
| **PAS** | Psychogeriatric Assessment Scale |
| **PROSPERO** | International Prospective Register of Systematic Reviews |
| **PsychInfo** | Database of abstracts of literature in the field of psychology. |
| **RUDAS** | The Rowland Universal Dementia Assessment Scale |
| **SIS** | Six item screener |
| **SPMSQ** | Short portable mental status questionnaire |
| **TICS** | Telephone Interview for Cognitive Status |
| **TRIP** | Turning Research Into Practice database |
| **TYM** | Test Your Memory |
| **VF-an** | Verbal Fluency - animals |
| **WHO** | World Health Organisation |

# 3. A rapid review of clinical practice guidelines relating to the use of brief cognitive assessments used as part of the process for diagnosing dementia within general practice

*MMSE: The gold standard of cognitive tests*

*Barrett, A & Burns, A (2014) Dementia revealed. What primary care needs to know. Hardwick CCG, NHS England, Department of Health, Royal College of General Practitioners [p.39]*

The focus of this chapter is the conduct and results of a rapid review of existing clinical practice guidelines (CPGs) relating to the use of BCAs as part of the process for diagnosing dementia within general practice. This rapid review emphasised the need for greater clarity in CPGs on this specific topic. This review also identified a necessity for more informed evidence on demand for CPGs, how they influence practice and how they can better support BCA use within this clinical population.

Context informing the review is initially discussed, followed by definition of the common terms used such as brief cognitive assessment, diagnosis in the general practice context, case finding and screening at the population level and the level of the individual patient. Search methods are described in detail, and results of the rapid review are presented relating to general practice, BCAs and CPGs included within the review. The discussion considers major and minor findings and their relevance to current practice. Recommendations for practice and research are made at the end of the discussion.

## 3.1. Why do a rapid review of clinical practice guidelines?

This review is intended to inform understanding of current clinical practice by identifying specific CPGs relating explicitly to the use of BCAs used as part of the process for diagnosing dementia within general practice.

### 3.1.1. What is a rapid review?

A rapid review is defined here as a type of knowledge synthesis in which standard systematic review processes are limited in order to produce a review within a restricted scope[121].

Rapid reviews are created through a transparent, rigorous and reproducible method of evidence synthesis that follows the key principles of systematic review. There are a number of strategies used to expedite the conduct of a rapid review, and the essential challenge is in maintaining robustness and transparency that are central to quality knowledge synthesis[122]. Rapid reviews are considered part of the same 'family' as systematic reviews. They should include a clear statement up front of the objectives of the review, predefine inclusion criteria, appraise the validity of findings, and present results and synthesis in a systematic way.

Whilst there is currently no generally agreed standardised approach to conducting rapid reviews[123], the key aspects of a rapid review listed above feature in much of the mainstream literature on rapid review methodology[121,122,124]. The three main ways that rapid reviews are made more rapid that systematic reviews are by: increasing resource above the usual systematic review team (e.g. a larger number of reviewers conduct parallel tasks such as screening studies for eligibility, abstracting data and assessing quality); applying shortcuts to the review, missing one or more systematic review steps; and automating processes within the review to fast-track e.g. data abstraction.

Within this rapid review, the review protocol was pre-registered on the international register for prospective registration of systematic reviews (PROSPERO) but was not published in an open access peer reviewed journal as this would have delayed the start of the review beyond reasonable timescales. Only papers published in English language were included within the review, as there was no additional resorce for translation services within the timeframe of the review. The literature search was curtailed from a standard 6 databases to 4 main databases, and searches were checked with an information specialist but designed and conducted by the single reviewer. All tasks were carried out by a single reviewer, meaning the standard independent double screening and abstraction that would take place in a systematic review was not possible within this rapid review. It is possible that these restrictions introduced inequalities including sampling, selection, and data accuracy biases. There is little known about the existence and extent of influence of these and other biases within the rapid review process, but these factors are considered within the Discussion section.

Detail of the rapid review processes followed here are provided in 3.4.1.

## 3.2. What do we mean by 'brief cognitive assessment'?

BCAs are one type of tool available for general practitioners (GPs) to use as part of their diagnostic assessment for someone with suspected dementia or other cognitive problems. A list of key BCAs currently available for use in contemporary primary care and general practice, developed and critically refined throughout this PhD project, is shown in Table 2.

Table 2.Comparison of commonly used assessment tools relative to cognitive domains

| Cognitive domain | MMSE | GPCOG | CDT | SIS | Mini-Cog | AMT | 6-CIT | ACE |
|---|---|---|---|---|---|---|---|---|
| **Memory** | | | | | | | | |
| **Semantic** | − | + | + | − | + | + | − | ++ |
| **STM** | + | ++ | − | + | + | + | ++ | +++ |
| **Remote** | − | − | − | − | − | + | − | ++ |
| **Visuospatial/ constructional praxis** | + | ++ | ++ | − | ++ | − | − | +++ |
| **Frontal/executive** | − | + | + | − | + | − | − | ++ |
| **Orientation** | +++ | + | − | + | − | ++ | ++ | +++ |
| **Attention/calculation** | ++ | + | + | − | + | ++ | ++ | ++ |
| **Language** | ++ | − | − | − | − | − | − | +++ |
| **Other aspects** | | | | | | | | |
| **Informant component** | − | + | − | − | − | − | − | − |
| **Equipment required** | Pen, paper & watch | Pen & paper | Pen & paper | − | Pen & paper | − | − | Pen, paper, watch & specialized pictures |

**−**, Not specifically tested; **+**, minimal assessment; **++**, moderate assessment; **+++**, relatively extensive assessment. Adapted from Woodford and George 2008[125]

### 3.2.1. Brief

The term "brief cognitive assessment" has been in common use since at least the 1980s [126-128], yet there is little definition within research literature of what this actually means. Undoubtedly subjective in nature, 'brief' seems in many cases to mean the test is perceived to take up to and including 10 minutes to administrate[64,129-131].

This apparently arbitrary window of time may at one point have represented the standard clinical appointment time with the family doctor; however, typical consultation time is a contested topic and evidence varies on the average length of general practice consultation. A large-scale 2002 study reported a UK general practice mean of 9.4 minutes (SD 4.7) to a mean of 15.6 minutes (SD 8.7) in Switzerland[132]. A more recent analysis of two practices in the UK suggest average consultation times of 15 minutes

or more[133], whereas a retrospective analysis of 100 million English GP consultations from 2007-2014 reported an average consultation time of 8.86 minutes (SD 8.86-8.87)[134]. In summary, there is no minimum GP consultation time in the UK and the Royal College of General Practitioners (RCGPs) states the usual length of time is 8-10 minutes[135].

Clearly, the length of time a test takes to conduct and its suitability for use in different populations is a political issue as much as an administrative one. Time within the consultation room is increasingly at a premium[136] and any test that claims to be 'brief' may attract greater support and uptake than those which claim otherwise or nothing at all, irrespective of the actual evidence base. It has also been suggested that a cognitive assessment's diagnostic accuracy should be balanced against its speed of administration, depending on the setting[137,138]. These challenges around the evaluation of test administration time are explored in more depth in the Discussion chapter.

Within this chapter and throughout the thesis, "brief" is used to refer to a test that takes up to and including 10 minutes to administrate within the general practice clinical setting, based upon empirical measures where possible, as well as (more commonly) the judgements of referenced review and study authors.

### 3.2.2. Cognitive assessment

Cognitive assessment concerns investigation of the higher cortical functions, namely memory, attention, orientation, language, executive function (planning activities), and praxis (sequencing of activities)[40]. Cognitive assessment is often positioned as an iterative process comprising various stages, in contrast to a single point valuation. This correctly reflects the clinical reality of cognitive assessment as a process, but poses a challenge when considering the purpose of a BCA is to provide a quick initial assessment of cognitive function, and the general practice population it may be used for, where (in the UK) consultations generally last around 8 minutes[132,133] and often multiple issues are discussed[134]. Often BCAs are used to make an initial assessment to aid decision making for further treatment, management or onwards referral to a specialist assessment unit such as a memory clinic. The diagnostic evaluation is not a straightforward concept, and relies on a number of aspects of testing such as test reliability (is it stable, does it get the same result time after time?), test validity (does

the test measure what it is supposed to measure?) and test accuracy (does the test correctly identify people with the condition and correctly rule out people without the condition?). These aspects are assessed through different methods, and some rely on dichotomous point estimates (e.g. correct/incorrect) which assume that a test is taken at one point in time. This sharply contrasts with the procedural reality of cognitive assessment within the GP clinic, and is another complexity of the process for identifying possible dementia in primary care.

There is no single "gold standard" cognitive assessment for identifying dementia [64,139,140] and whilst there is broad agreement amongst numerous guidelines on dementia identification within family practice, variation also exists around the different tools recommended.

### 3.2.3. Single tests vs. batteries

There are large number of cognitive assessments available, many calling themselves 'brief'. These assessments can be categorised further in a number of ways. One of the most common distinctions is whether the assessment is a single test (such as the Clock Drawing Test (CDT), where the individual is asked to draw a clock face and mark a specific time upon it), or a battery such as the Abbreviated Mental Test Score (AMTS). Sometimes these are referred to as 'single domain' tests which address one cognitive domain such as visual-spatial, in the case of the CDT, or 'multi-domain' tests which address several cognitive domains such as episodic memory, semantic memory, short-term memory, attention, and orientation such as the AMTS [141].

Table 3 List of BCAs suitable for use in general practice

| BCA name + validated study reference | Informant section (yes/no) | Suitable for particular groups |
|---|---|---|
| GPCOG[142] | Yes | Patients with informant (optional) |
| CDT[143] | No | Patients who can hold a pencil |
| Mini-Cog[144] | No | Non-native speakers and people with low levels of education |
| IQCODE[145] | Yes | Non-Caucasian populations |
| SPMSQ[146] | No | Elderly patients |
| AMTS[146] | No | Elderly patients |
| 6-CIT[147] | Yes | - |
| RUDAS[148] | No | Culturally and linguistically diverse populations |
| MIS[149] | No | - |
| TICS[150] | No | Geographically remote or less mobile populations |

GPCOG, General Practitioner Assessment of Cognition; CDT, Clock Drawing Test; IQCODE, Informant Questionnaire on Cognitive Decline in the Elderly; SPMSQ, Short Portable Mental Status Questionnaire; AMTS, Abbreviated Mental Test Score; RUDAS, Rowland Universal Dementia Assessment Scale; 6-CIT, Six Item Cognitive Impairment Test; MIS, Memory Impairment Screen; TICS, The Telephone Interview for Cognitive Status.

Some have argued for reliance on assessing short term memory and executive function as these are commonly associated with Alzheimer's disease as the most frequently occurring dementia subtype[117]. Others[125] point to the benefit of single tests - used within appropriate populations - as generally simpler and quicker to administer than multi-domain assessments, thus improving clinician acceptance of the test as demonstrated with the MMSE[151].

Key distinctions between these different types of assessments are the different cognitive areas or domains that they assess (e.g. visual-spatial, episodic memory, orientation), and the administrative nuances which these distinctions introduce. For example, the 'Verbal Fluency-animals' (VF-an) test assesses a person's ability to name pictures of different animals presented to them, and is scored on time taken and accuracy of the naming. This assesses language domains, but clearly also relies on functional hearing, speech and eyesight as well as language skills – aspects that could easily be compromised and not necessarily recognised by the healthcare professional carrying out the assessment. Measures incorporating behavioural and functional assessment are increasingly popular as these factors are recognised as early markers for timely dementia diagnosis[39].

The length of time that single- and multi-domain tests take to carry out is a decisive factor in their relevance for use in primary care, and tools such as the 7 Minute Screen (taking around 12 minutes [152]), the Leganés Cognitive Test[153] and the Addenbroke's Cognitive Examination-Revised (ACE-R)[154] are considered too long for standard general practice consultation lengths. There is also little evidence that tests with longer administration times perform better[139].

### 3.2.4. Other aspects of cognitive function

Cognitive assessments can consist of tools completed with or by the individual, and some cognitive assessments also incorporate an assessment to be completed by a close friend, partner or family member (often referred to as an 'informant'). These informant assessments are generally an addition to the main cognitive assessment, and can either sit alongside the main assessment as an additional tool to increase the test's effectiveness, or used as a 'stepped' assessment where, if a certain threshold is reached on the assessment of the individual, the informant assessment is carried out and the scores combined to create a distinct 'staged' assessment tool. Examples of informant assessments are the 'Informant' section of the General Practitioner Cognitive Assessment tool (GPCOG)[142] and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)[103].

### 3.2.5. Why settings matter

The accuracy of tests of cognitive function has been explored across a range of settings[42,102,116,155-157], and some have also looked at the strength of evidence for these accuracy scores[91,99,117,156,158].

Whilst the setting in which these tests are validated has received less scrutiny within the research literature, the distinction between the original validation population and the population within which the tests are ultimately used is a vital and variable aspect of test evaluation. In many countries there is an increased impetus to proactively identify dementia earlier in the diagnostic pathway. This is in order to rule out potentially treatable alternative and confounding conditions (such as depression, urinary tract infections or thyroid disorders), alleviate symptoms and access services as soon as possible to increase time for future planning, reduce cost and diminish care burdens on the patient and the people around them. This is reinforced by policy within the UK healthcare context[94] as well as across European systems such as in Spain,

Portugal, Ireland, France, the Netherlands and Belgium[39], where general practitioners or family doctors are often the first place where people present with concerns about their own or others' possible dementia. In contrast, secondary care has a role in making a aetiological diagnosis, identifying subtypes and stratifying patients by severity of symptoms - although in reality these services overlap considerably[88]. More detailed consideration of these methodological issues are presented in the Discussion chapter of this thesis, including worked examples of how prevalence influences false discovery rates (numbers of false positives and false negatives).

When choosing between tests, it is clearly important to take into account the population in which the test is intended to be used. Guidance on the suitability of different tests for different populations is often unclear[159], and the challenges underpinning this obscurity will be explored in the next section.

### 3.2.6. What are CPGs?

Guidelines are a general rule, principle or piece of advice[160]. Some guidelines are referred to as strategies, plans or frameworks. The rapid review presented here deals specifically with CPGs, defined as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances."[161].

CPGs are often produced within the healthcare context in order to recommend how health care professionals can best care for people with a certain condition. These guidelines cover many different aspects of a condition and may incorporate recommendations for giving information and advice, prevention, diagnosis, treatment and longer-term management[162]. CPGs have been published for many years with common aims to improve the selection, use and interpretation of BCAs within the intended populations, and can be an important tool for the management of dementia although there is mixed evidence of how they are valued by healthcare professionals[39].

### 3.2.7. Who produces and issues CPGs?

There has been a significant growth in the creation of national general guidelines for general practitioners managing dementia in recent years. Before 2003 there were no guidelines for identifying dementia in general practice within some of the major European healthcare systems (Ireland, Italy and Portugal)[39]. Currently Ireland has established guidelines[163], whereas Italy[164] and Portugal[165] are developing them as

part of their national dementia strategies. Across Europe, there are a large number of countries with dementia strategies that incorporate guidelines for clinical practice (UK – including separate guidelines for England[94], Northern Ireland[166], Scotland[167] and Wales[168]; Ireland[163]; Norway; Finland; Denmark; Belgium; Luxemburg; Czech Republic; Austria; Slovenia; Switzerland; Greece; Sardinia[169] and outside of Europe countries such as Israel[170], Australia[111], the USA[62], Canada[110] - all have dementia-specific general guidelines.

Doctors often produce guidelines for other doctors[39,171], and whilst this may increase usability amongst clinicians this may limit their accessibility for other guideline users involved in wider (social, cultural) aspects of care. Less formally, local general practices may have their own implicit guidelines for dementia assessment which are not published but may have the greatest impact on what brief cognitive assessment procedures are used at the local level.

It is unclear how many unequivocal formal CPGs there are, guiding the use of BCAs in primary care. This rapid review was therefore conducted in order to systematically identify CPGs relating to the use of the BCAs used as part of the process for identifying dementia in general practice.

## 3.3. Methods

CPGs are defined here as “systematically developed statements to assist practitioner decisions about appropriate health care for specific clinical circumstances.”[172]

This is a rapid review and not a full systematic review. It has had restrictions placed upon the conduct in terms of timeframe for implementation, limited resources and conduct by a single researcher. However, the review was carried out systematically and has many of elements of best practice systematic review methods: the question was specified *a priori*, the protocol was pre-registered on the international register for prospective registration of systematic reviews (PROSPERO reference CRD42019139159), the search criteria were explicitly stated, checked with an information specialist and published in detail, alongside pre-defined inclusion and exclusion criteria, a pre-defined data extraction form was created, and the discussion provides limitations of included studies and the review process[124].

*Searches*

Searches were structured by the author to identify the best and most relevant evidence likely to help answer the research question within the parameters of the method. This search framework was checked with two expert and experienced information specialists, with feedback incorporated *a priori* into the final searches run.

Terms were based on MeSH headings and were practice guideline; dementia/di [Diagnosis]; and general practice/ or family practice.

Databases searched were EMBASE, MEDLINE, PsychInfo and TRIP. Within the TRIP database, the 'Guidelines' filter was applied to identify the most relevant evidence specifically related to guidelines. There was no date filter used.

PROSPERO was searched for any systematic reviews already in motion on this specific topic. There were no systematic reviews found to be published or registered on PROSPERO in this particular area.

*BCAs*

Whilst there is no internationally-accepted single definition of 'brief cognitive assessment', the term understood for the purposes of this review acknowledges the iterative nature of the process, drawing on a number of observations within a set ('brief') timescale of 10 minutes and their use in general practice to make an initial indicative assessment for further investigation within more specialist settings such as memory clinics.

*General practice*

General practice as a term is sometimes mistakenly used interchangeably with primary care, but is more accurately defined as first line clinical practice delivered by general medical clinicians and health care professionals trained in general medicine. In non-UK settings this is sometimes referred to as family practice.

Items were only be included if they clearly and explicitly referred to themselves as CPGs. Items also related explicitly to general practice or family practice, and explicitly referred exclusively to the use of BCAs used as part of the process for identifying dementia in this setting. If an identified CPG was superseded entirely by a more recent CPG issued by the same body, then the older CPG was excluded. The exception to

this would be where the older CPG contained general practice-relevant guidance, and the updated CPG referred to the older document.

*Language*

Only guidelines published in English were included.

*Primary outcome*

The single outcome of interest for this rapid review is the CPG produced explicitly for general practice relating specifically to the use of BCAs for identifying dementia in general practice.

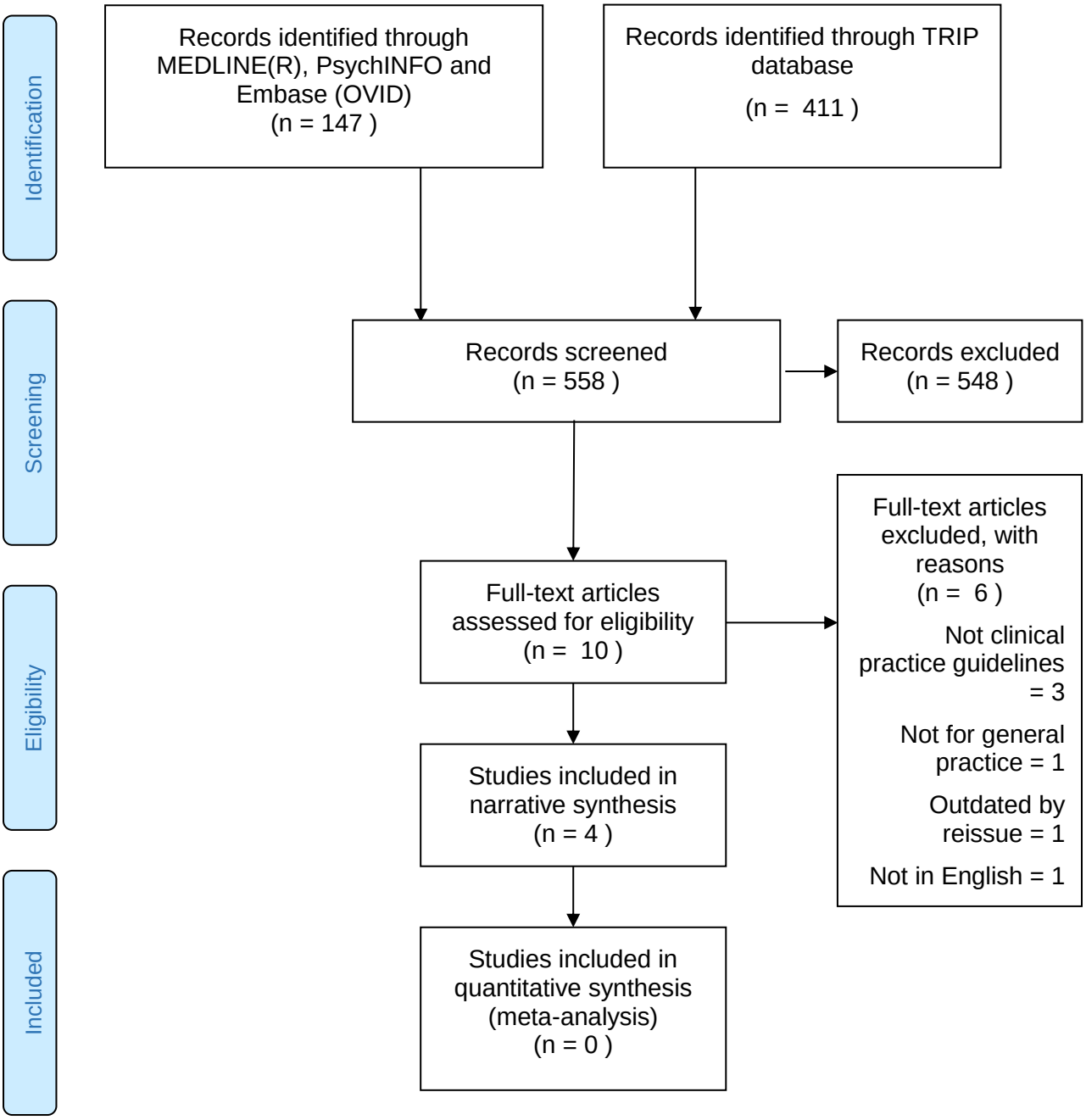
## 3.4. Results

From 558 records identified through database searches, 10 hits were identified for inclusion at the title/abstract screening stage. Screening was conducted by one researcher (HH).

At full text screening, 6 papers were excluded; three on the basis that they did not directly report CPGs, one because it related to psychologists rather than general practice, and one issued by the Dutch College of General Practitioners in 2004 because it was outdated by a more recent guideline issued by the same organisation in 2012, with no additional information useful for the review. The later paper from the Dutch College of General Practitioners published in 2012 was excluded as it was not available in English. Screening was conducted by one researcher (HH). The review flow diagram is shown in

Figure 2.

Figure 2. PRISMA flow diagram for rapid review of CPGs



Papers excluded at the full text screening stage are shown in Table 4, along with the reasons they were excluded.

Table 4. Papers excluded at the full text screening stage, with reasons for exclusion

| Full reference | Reason for exclusion at full text screening stage |
|---|---|
| Association AP. Guidelines for the evaluation of dementia and age-related cognitive change. The American psychologist. 2012;67(1):1. | Related to psychologists rather than general practice |
| Clarfield A. Canadian Consensus Conference on the Assessment of Dementia reports. Canadian Family Physician. 1997;43:1343. | Does not directly report clinical practice guidelines |
| Dyer SM, Laver K, Pond CD, Cumming RG, Whitehead C, Crotty M. Clinical practice guidelines and principles of care for people with dementia in Australia. Australian family physician. 2016;45(12):884. | Does not directly report clinical practice guidelines |
| Wilcock J, Iliffe S, Turner S, et al. Concordance with clinical practice guidelines for dementia in general practice. Aging and Mental Health. 2009;13(2):155-161. | Does not directly report clinical practice guidelines |
| Boomsma L, Boukes F, Wind A, Assendelft W. Summary of the practice guideline'Dementia'(second revision) from the Dutch College of General Practitioners. Nederlands tijdschrift voor geneeskunde. 2004;148(24):1191-1197. | Outdated by a more recent guideline issued by the same organisation in 2012 |
| Luning-Koster M, Perry M, van Charante Moll E, Vernooij-Dassen M, Wiersma T, Burgers JS. Summary of Dutch College of General Practitioners'(NHG) practice guideline'Dementia'. Nederlands tijdschrift voor geneeskunde. 2012;156(49):A5323-A5323. | Not in English |

*Overview of results*

Following systematic searches of four international databases, four CPGs were identified that were produced *explicitly* for an audience including general practice relating specifically to the use of BCAs for identifying dementia in general practice[110,173-175]. The characteristics of these guidelines are shown in Table 5.

Table 5. Summary of characteristics for included guidelines

| Full guideline reference | Year of issue | Self-refers as CPG? | Country related to | All guidelines relevant for general practice? | BCAs covered | Detail given |
|---|---|---|---|---|---|---|
| National Institute for Health and Care Excellence Dementia: Assessment, management and support for people living with dementia and their carers: National Institute for Health and Care Excellence London; 2018. | 2018 | Yes | United Kingdom | No, some relate to specialist assessment, community care, hospitals and care homes. | 10-CS<br>6CIT<br>MIS<br>TYM<br>IQCODE | In the conduct of cognitive tests, a validated brief structured cognitive instrument should be used such as those in the examples given. Dementia should not be ruled out solely because the person has a normal score on a cognitive instrument. |
| Guideline Adaptation Committee GA. Clinical practice guidelines and principles of care for people with dementia. Sydney: Guideline Adaptation Committee. 2016. | 2016 | Yes | Australia | No, some relate to neuroimaging. | 3MS<br>MMSE<br>ADAS-Cog<br>GPCOG<br>PAS<br>RUDAS<br>KICA-Cog<br>MoCA<br>FAB<br>ACE-*III* | Based on UK NICE guidelines, adapted for Australian context.<br>List of BCAs *adapted from the Dementia Outcomes Measurement Suite [DOMS:* www.dementia-assessment.com.au*]* |
| Gauthier S, Patterson C, Chertkow H, et al. Recommendations of the 4th Canadian Consensus Conference on the Diagnosis and Treatment of Dementia (CCCDTD4). Canadian Geriatrics Journal. 2012;15(4):120. | 2012 | Yes | Canada | No, but symptomatic treatments are relevant to all practitioners. | Referral to previous guidelines (see Kennedy 2006 below). | Only treatments covered, no guidelines for general/family practice |
| Kennedy HDP. 3rd Canadian Consensus Conference on Diagnosis and Treatment of Dementia. 2006. | 2006 | Yes | Canada | No, but symptomatic treatments are relevant to all practitioners. | GPCOG<br>MoCA<br>DemTect<br>BNA<br>7MS | |

Brief Cognitive Assessments (BCS); Modified Mini Mental Exam (3MS); Mini Mental State Exam (MMSE); The Alzheimer's Disease Assessment Scale - Cognition (ADAS-Cog); General Practitioner Assessment of Cognition (GPCOG); Psychogeriatric Assessment Scale (PAS); Rowland Universal Dementia Assessment Scale (RUDAS); Kimberley Indigenous Cognitive Assessment (KICA-Cog); Montreal Cognitive Assessment (MoCA); Frontal Assessment Battery (FAB); Addenbrooke's Cognitive Examination (ACE-R now replaced by ACE-III); the 10-point cognitive screener (10-CS); the 6-item cognitive impairment test (6CIT); the 6-item screener; the Memory Impairment Screen (MIS); the Mini-Cog; Test Your Memory (TYM); Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE), Behavioural Neurology Assessment (BNA), 7 Minute Screen (7MS)

.

All CPGs were single national guidelines. One was from the UK[173], one from Australia[174] and two from Canada[110,175]. All were published in the last 13 years, with the UK CPG published by the National Institute for Health and Care Excellence (NICE) in 2018 being the most up to date. The Canadian CPGs[110,175] produced by the Canadian Consensus Conferences on the Diagnosis and Treatment of Dementia (CCCDTD) combines CPGs with consensus statements, resulting from a consensus conference with 80% or more of voting attendees supporting agreed statements. The Australian CPG[174] produced by the Cognitive Decline Partnership Centre (CDPC) combines CPGs with 10 Principles of Care, developed by the Social Care Institute for Excellence in the UK based on consultation with members of the public[176]. The UK CPG addresses diagnosis, alongside management and support for people living with dementia and their carers[173].

All included CPGs were created for an audience wider than general practice, including hospitals, community care, specialist assessment, and care homes. This means there are a number of guidelines within all three CPGs that are not applicable to general practice.

The Canadian CPGs[110,175] include guidelines on neuroimaging, and the most recent Canadian CPG[110] covers new diagnostic criteria, neuroimaging and liquid biomarkers. This CPG also contains updated guidance on symptomatic treatments. The Australian CPG[174] includes guidelines on diagnosis including neuroimaging, treatment, and social support for people with dementia and carers. The UK CPG[173] covers guidelines on diagnosis in non-specialist and specialist settings with guidance including neuroimaging and detailed neuropsychological evaluation, shared decision making, management, treatment and care coordination including training dealing with comorbidities.

*Guidance in CPGs relating to diagnosis of dementia in general practice*

The UK CPG recommends that at the initial assessment, a patient history is taken from the individual and *ideally* from someone who knows the person well. If dementia is still suspected after this assessment, then the patient should have a physical examination, alongside appropriate blood and urine tests to rule out other causes of cognitive problems, and cognitive tests should be conducted.

The Australian CPG are based closely on the UK CPG, but the guideline committee noted some important differences related to indigenous communities in Australia. They recognise a higher prevalence of dementia in Indigenous Australians, and the possibility that dementia may present at an earlier age in some communities. There may be different perceptions of alterations in cognition in some cultures, leading to individuals and their carers seeking diagnosis at a later stage of the dementia progression.

The most recent 2012 Canadian CPG does not include any recommendations specific to diagnosis in general practice or primary care, and the authors state that general clinicians are unlikely to be substantially affected by the guidelines. Previous Canadian CPGs[175] produced in 2006 make recommendations for BCAs suitable for general practice, but do not give guidance on diagnostic procedure in general practice.

*Guidance in CPGs relating to specific BCAs to use as part of the diagnosis of dementia in general practice*

All BCAs included within the CPGs are shown in Table 6. The UK CPG is carefully worded to say that no single test is recommended, but when cognitive testing is carried out in a non-specialist setting, a validated BCA should be used such as those shown in Table 6. The UK guidance does clearly state that dementia should not be ruled out because the person has a normal score on a BCA.

The Australian CPG is more directive, with instructions that professionals conducting cognitive assessment should take into account all other factors known to affect performance such as age, educational level, non-English speaking status, prior level of functioning, aphasia, visual or hearing impairments, psychiatric illness or physical/neurological problems when interpreting results. This guidance also recommends formal neuropsychological testing in cases where a dementia diagnosis is uncertain.

The 2006 Canadian CPG[175] reports a systematic assessment of the strength of evidence for recommending the BCA list in Table 6, and concludes that there is insufficient evidence for recommending one over another, as replication studies are needed.

Table 6.CPG recommendations on BCAs to use as part of the diagnostic process for identifying dementia in general practice

| BCAs listed in CPG | UK[173] | Australia[111] | Canada[175] | Notes on the recommendation |
|---|---|---|---|---|
| The 10-point cognitive screener (10-CS) | | | | Do not rule out dementia solely because the person has a normal score on a cognitive instrument. |
| The 6-item cognitive impairment test (6CIT) | | | | |
| The 6-item screener | | | | |
| The Memory Impairment Screen (MIS) | | | | |
| The Mini-Cog | | | | |
| Test Your Memory (TYM) | | | | |
| Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) | | | | When taking a history from someone who knows the person with suspected dementia, consider supplementing this with a structured instrument such as IQCODE or the FAQ |
| Functional Activities Questionnaire (FAQ). | | | | |
| Kimberley Indigenous Cognitive Assessment (KICA-Cog) or KICA-Screen tool | | | | Recommended for use with remote living Indigenous Australians for whom the use of alternative cognitive assessment tools is not considered appropriate. |
| The Rowland Universal Dementia Assessment Scale (RUDAS) | | | | Should be considered for assessing cognition in culturally and linguistically diverse (CALD) populations. |
| Modified Mini Mental Exam (3MS) | | | | Recommended for assessment of cognitive function based on the appraisal in the Dementia Outcomes Measurement Suite *[DOMS: www.dementia-assessment.com.au]* |
| Mini Mental State Exam (MMSE) | | | | |
| The Alzheimer's Disease Assessment Scale - Cognition (ADAS-Cog) | | | | |
| Frontal Assessment Battery (FAB) | | | | |
| Psychogeriatric Assessment Scale (PAS) | | | | |
| Addenbrooke's Cognitive Examination (ACE-R now replaced by ACE-III | | | | |
| General Practitioner Assessment of Cognition (GPCOG) | | | | |
| Montreal Cognitive Assessment (MoCA | | | | |
| DemTect, | | | | |
| Behavioural Neurology Assessment (BNA) | | | | |
| 7 Minute Screen (7MS) | | | | |

Table 6 shows that only two BCAs – MoCA and GPCOG – feature in more than one guideline within the CPGs of Canada and Australia, and no tests feature in all three guidelines.

It is particularly interesting that the CPGs of UK and Australia do not recommend any of the BCAs, given the Australia CPG is explicitly based upon the UK CPG, although adapted for Australian populations.

Both CPGs from Canada and Australia make reference to the importance of recognising indigenous communities within their guidance, but only the Australian CPG includes BCAs recommended for use with indigenous and culturally and linguistically diverse populations.

None of the included guidelines give further information on the BCAs listed in Table 5, in terms of, for example: time to administrate; cost; resource required; patient and clinician acceptability.

## 3.5. Discussion/ Conclusion

*Small number of CPGs identified*

It was a surprising finding that so few CPGs were identified within the searches presented here. This may be for a number of reasons. The searches may not have been sensitive enough, and may have missed guidelines that did not explicitly define themselves as CPGs. However, as identifying specific CPGs was the clear focus of this rapid review, it was a logical requirement to place on identified evidence that they must explicitly self-identify as a CPG.

It may be that existing CPGs are not indexed within the databases searched, as grey literature sources and websites were not included within searches for this review. They may instead be published within national government or other organisational websites, or even possibly not widely published or available beyond the clinical boundaries within which they are widely used. This would, however, be surprising, given the breadth of coverage within EMBASE, MEDLINE, PsychInfo and TRIP and the fact that CPGs are generally produced with an explicit aim to increase coverage and take-up amongst their target population. It is therefore possible, but unlikely, that existing CPGs may have been missing from these databases. Placing an English language restriction on this review did mean that CPGs not published in English were excluded, but there were only two CPGs produced by the Dutch College of General Practitioners

in 2012[177] and 2004[178] that were excluded purely because they were not written in English.

The most compelling explanation that merits some consideration is that the CPGs do not yet exist, so could not be identified within this rapid review.. As shown in Table 6, out of 21 BCAs listed within the guidelines, only two are common across two guidelines – and none of the BCAs feature in all three CPGs. This is a significant result, particularly given the currency of these CPGs, all produced within the last 7 years, and the fact that the Australian CPG was explicitly based upon the UK CPG yet shared none of the recommended BCAs.

*Lack of guidance on CPG tailored to population*

Another notable gap within identified CPGs is the scarcity of guidance on tailoring CPGs for specific populations. Given the national reach of these three CPGs, only the Australian CPG contains clear guidance on the suitability of BCAs for indigenous Australians, urban and rural populations and cultural and linguistically diverse populations.

Neither the Canadian nor UK CPGs make explicit reference to BCAs for different populations, nor do they make recommendations on the diagnostic process tailored for diverse groups. The UK CPG instead places emphasis on treating patients as individuals and respecting their culture and identity as part of that awareness. This overlooks the reality that many BCAs contain innate demographic and cultural bias within their scoring systems concerning learning difficulties, disabilities (such as sensory or physical impairments), level of education, sex and cultural background[125,148,179].

*Lack of common BCAs within guidance*

The absence of common BCAs recommended across the three identified CPGs warrants further exploration. This lack of shared guidance, or even shared recognition of BCAs suitable for use as part of the process for identifying dementia in general practice, indicates a deeper problem in identifying and recommending suitable tools for diagnosis within a general practice setting.

There are 21 BCAs recommended within the three CPGs identified within this rapid review. Of those 21 BCAs, 2 (MoCA and GPCOG) are shared between two CPGs (Australia and Canada). Of these 2 BCAs, the MoCA takes between 10-15

minutes[42,64,180] to administer which is at the limit of the entire 10 minutes allocated for most UK and European general practice consultations[132-134]. It is therefore questionable whether any BCA of this length would be suitable for use in general practice, where time and clinician resource is at a premium and GP preference is for shorter tests[133,136,181,182].

The lack of uniformity in BCAs featured or recommended within the three identified CPGs is a strong indicator that there remains a gap in knowledge about which BCAs are most suitable for use in a general practice setting and population. This may be because the suitability of BCAs within general practice is not fully understood, or it may be due to a fundamental lack of suitable BCAs fit for the purpose of aiding diagnosis of dementia within a primary care, non-specialist setting. The solution probably lies in both these issues; understanding what it is that people value in BCAs, above and beyond basic test accuracy but including aspects such as acceptability to the patient and to the clinician, accessibility, cost, time to administer and applicability for different populations and different settings such as care homes, and in the community. These fundamental features of test design, choice and advocacy are crucial to understand and apply clearly within CPGs, yet this rapid review has highlighted a significant dearth in uniform, unambiguous guidance relating specifically to BCA use as part of the process for identifying dementia in general practice.

These issues are discussed in further detail within the rest of this thesis; in the systematic overview of the diagnostic accuracy of BCAs for identifying dementia in general practice (Chapter 4), the systematic review directly comparing the diagnostic accuracy of MMSE and GPCOG (Chapter 5), and within the main thesis Discussion (Chapter 7).

*Recommendations for practice*

The primary recommendation from this rapid review for practice is that greater clarity is needed in CPGs on this specific topic. There were only three CPGs that were identified within this paid review, all from European style healthcare systems (UK, Australia and Canada) with broadly similar general practice populations, excepting indigenous communities.

The lack of diversity and breadth of CPGs from other countries is notable, although it should be remembered that CPGs not available in English language were excluded.

With this caveat, the apparent limited range of CPGs is something to be explored further – not only in order to aid clinicians and general practitioners from other countries, but in order to tackle the current gaps in knowledge and understanding of a complex process. General practice is regularly highlighted as a 'choke point' for dementia diagnosis[138,183-187] yet the challenges presented that are specific to this setting and population mean that tailored guidance is acutely needed within this field.

Greater concordance amongst CPGs will only improve adoption, help clinical practice and improve understanding of the place of BCAs within the diagnostic assessment in a non-specialist setting. This may not be possible, however, until we have a better understanding of which BCAs are best for these requirements, and how BCAs are valued in general practice. These far more difficult questions need to be addressed first, in order to help bring consistency and standardisation to CPGs.

*Recommendations for research*

Currently there is insufficient validation and replication of BCAs for use in general practice. These pieces of work are urgently needed in order to critically assess BCA suitability for different general practice populations, variations in setting, and performance in identifying differing levels of severity of cognitive impairment.

Clear, robust assessment of the validity of BCAs in general practice settings and populations is needed, particularly assessments taking account of diversity and variation at the population level, not just at the individual level as acknowledged within the UK CPG[173].

There is a strong drive to understand the needs of general practice relating to CPGs on BCAs. Clearly in order to response to need, maximise practitioner adoption and ensure that guidelines are fit for purpose there remains much to be understood about CPG usage in general practice, how BCAs are selected and used within general practice currently, how GPs decide on BCAs within a clinical consultation, how that relates to the patient and how CPGs influence GP, commissioner and patient behaviour in choosing and using BCAs.

Once there is greater understanding of the critical features of BCAs for general practice use, it is highly likely that CPGs will be far more available, coherent and cohesive for the intended general practice audience.

## Glossary

| | |
|---|---|
| **3MS** | Modified Mini Mental Exam |
| **6-CIT** | 6 Item cognitive impairment test |
| **7MS** | 7 minute screen |
| **10-CS** | 10 point Cognitive Screener |
| **ACE** | Addenbrooke's Cognitive Evaluation |
| **ACE-R** | Addenbrooke's Cognitive Evaluation - revised |
| **ACE-III** | Addenbrooke's Cognitive Evaluation – version three |
| **ADAS-Cog** | The Alzheimer's Disease Assessment Scale - Cognition |
| **ADI** | Alzheimer's disease International |
| **AGECAT** | Automated Geriatric Examination for Computer Assisted Taxonomy |
| **AMT(S)** | Abbreviated Mental Test Score |
| **BIMCT** | Blessed Information-Memory-Concentration Test |
| **BNT** | Behavioural Neurology Assessment |
| **CAMCOG** | The Cambridge Cognitive Examination |
| **CAMDEX** | Cambridge Mental Disorders of the Elderly Examination |
| **CASI** | Cognitive Abilities Screening Instrument |
| **CCCDTD** | Canadian Consensus Conferences on the Diagnosis and Treatment of Dementia |
| **CCG** | Clinical Commissioning Group |
| **CDPC** | Cognitive Decline Partnership Centre |
| **CDR** | Clinical Dementia Rating |
| **CDT** | clock drawing test |
| **CERAD** | Consortium to Establish a Registry for Alzheimer's Disease |
| **CPG** | Clinical Practice Guideline |
| **DSM-III** | The Diagnostic and Statistical Manual of Mental Disorders (version 3) |
| **DSM-III-R** | The Diagnostic and Statistical Manual of Mental Disorders (version 3 revised) |
| **DSM-IV** | The Diagnostic and Statistical Manual of Mental Disorders (version 4) |
| **DSM-IV-R** | The Diagnostic and Statistical Manual of Mental Disorders (version 4 revised) |
| **EMBASE** | Excerpta Medica dataBASE |
| **FAB** | Frontal Assessment Battery |
| **FAQ** | Functional Activities Questionnaire |
| **FCSRT** | Free and Cued Selective Reminding Test |
| **GMS-AGECAT** | Geriatric Mental State Schedule - Automated Geriatric Examination for Computer Assisted Taxonomy |
| **GP** | General Practitioner |
| **GPCOG** | The General Practitioner assessment of Cognition |
| **ICD-10** | International Classification of Disease – version 10 |
| **IQCODE** | Informant Questionnaire on Cognitive Decline in the Elderly |
| **IPA/WHO criteria** | International Psychogeriatric Association/World Health organisation criteria |
| **KICA-Cog** | Kimberley Indigenous Cognitive Assessment |

| | |
|---|---|
| **KICA-Screen** | Kimberley Indigenous Cognitive screening tool |
| **MCI** | Mild Cognitive Impairment |
| **MEDLINE** | Medical Literature Analysis and Retrieval System Online |
| **MeSH** | Medical Subject Headings |
| **MIS** | Memory Impairment Screen |
| **MMSE** | Mini mental state examination |
| **MoCA** | Montreal Cognitive Assessment |
| **NHS** | National Health Service (UK) |
| **NICE** | National Institute for Health and Care Excellence (UK) |
| **NINDCS-ADRDA** | National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association |
| **NINCDS-AIREN** | National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l'Enseignement en Neurosciences |
| **NINCDR-CERAD** | National Institute of Neurological Disorders Clinical Dementia Rating -Consortium to Establish a Registry for Alzheimer’s Disease |
| **PAS** | Psychogeriatric Assessment Scale |
| **PCL** | Prueba cognitive de leganes |
| **PROSPERO** | International Prospective Register of Systematic Reviews |
| **PsychInfo** | Database of abstracts of literature in the field of psychology. |
| **RUDAS** | The Rowland Universal Dementia Assessment Scale |
| **SASSI** | Short and Sweet Screening Instrument |
| **SIS** | Six item screener |
| **SPMSQ** | Short portable mental status questionnaire |
| **TICS** | Telephone Interview for Cognitive Status |
| **TRIP** | Turning Research Into Practice database |
| **TYM** | Test Your Memory |
| **VF-an** | Verbal Fluency - animals |
| **WHO** | World Health Organisation |

# 4. An overview of systematic reviews summarising the accuracy of brief cognitive assessments for identifying dementia in primary care

*MMSE: The gold standard of cognitive tests*

*Barrett, A & Burns, A (2014) Dementia revealed. What primary care needs to know. Hardwick CCG, NHS England, Department of Health, Royal College of General Practitioners [p.39]*

In order to understand the spread of existing systematic review evidence on the accuracy of brief cognitive assessments used to identify dementia in primary care (and specifically in general practice), an overview of systematic reviews was carried out. Through the conduct of this review, it was hypothesised that systematic evaluation of the evidence would elicit a clear and definitive answer to which brief cognitive assessment is most accurate for use in primary care as part of the toolkit for identifying dementia. This chapter addresses methodological decisions made, key topic criteria, results from the overview and reflections of these methods and findings in the context of implications for clinical practice, research and my wider thesis. The overview aimed to answer the question: which brief cognitive assessment has the best diagnostic accuracy? This question could not be clearly answered within this overview, and the reasons for this are discussed within this chapter.

## 4.1. What is an overview?

Overviews are syntheses of systematic review evidence and are known by a variety of different names, all potentially reflecting different aspects and aims of the syntheses. Terms used include: 'overview of systematic reviews' (often shortened to 'overview') umbrella review; meta-review; (systematic) review of (systematic) reviews; synthesis of systematic reviews; and summary of systematic reviews. The common feature of the methods associated with all of these terms is the fundamental process of synthesising evidence which is derived, often exclusively, from systematic reviews. The systematic review forms the primary 'unit of analysis' and is the basis upon which an overview is built[188].

The term 'overview of systematic reviews' (often shortened to 'overview') appears to have gained the most widespread acceptance[189,190]. This is the term used by

Cochrane to describe a review of systematic reviews published in the Cochrane Library[191]. Here the term 'overview' is used within this thesis to describe summaries of systematic review evidence, in line with the most commonly-used terminology.

### 4.1.1. Why an overview may add value, and the challenges of overviews

As with systematic reviews, a primary consideration for producing an overview is whether there is sufficient justification for carrying it out[192]. This means the aims and objectives of the overview should be clearly understood and pre-specified before synthesis begins, and the methodological rigour employed within systematic reviews is (at least) as important when conducting an overview[193]. Decisions in healthcare should be based on the best evidence available rather than the most prominent, most recent or largest study. Overviews and systematic reviews may help to avoid research waste, duplication of effort, and allow anyone to view the best available evidence presented in a systematic way with quality of available evidence clearly presented to allow comparison across different studies[189,194].

Other rationales often given are that overviews can form a 'friendly front end' to existing evidence in order to allow more detailed investigation of systematic review evidence[189]. It has also been suggested that evidence can be gathered more quickly and easily from systematic reviews, and so an overview may be a more efficient form of evidence synthesis to address a specific research question[195] - although this has not been supported by the evidence so far.

There are different types of overviews that may be conducted including overviews addressing questions of interventions, prognosis or prevalence, diagnostic test accuracy, or qualitative views or experiences. The primary route to addressing an overview's aims and objectives is commonly through the structured clinical research question. As an example, when addressing existing evidence of diagnostic accuracy an overview may investigate the effect of a single test on a defined population, or may summarise evidence of the effect of a number of tests on a defined population. In situations where the test or population is ill-defined or cannot be fully known, an overview is highly unsuitable as problems inherent with such an approach in systematic reviews (such as avoidable heterogeneity and lack of generalisability) are magnified when abstracted to the level of an overview. In summary, the ideal overview is formed from a clear clinical question which results in a convergence and

strengthening of evidence on a given topic. How successful this overview was in matching these ideals will be discussed in this chapter, with specific methodological challenges addressed in Chapter 7.

### 4.1.2. Cochrane & non-Cochrane systematic reviews

There are two likely approaches to the *type* of systematic review evidence considered within an overview: one which *only* includes Cochrane reviews, versus a more inclusive approach of accepting all systematic reviews that meet the inclusion criteria.

This could be distilled into two strategies: the 'best evidence' approach, where the review team assumes that – where compared to other systematic reviews - Cochrane reviews represent the best available systematic review evidence on a topic with the highest quality standards, most up-to-date evidence and greatest rigour in analysis[196,197]. Alternatively the reviewer can follow the more usual 'comprehensive' approach where all evidence is sifted for eligibility and refined with large amounts of data sifted out at an early stage.

Whilst the best evidence approach had attractions in terms of greatest efficiency for resources and time as searches could be limited to the Cochrane Library, after some discussion with the review team it was decided to take the comprehensive approach to identifying evidence. It was concluded that there are many reasons that systematic review evidence is produced outside of the Cochrane system, including lack of time, inefficiency of editorial processes and bureaucracy in review production. Whilst systematic review quality is highly mixed and many systematic reviews are unnecessary, misleading and/or conflicted[193], the reliable supremacy of Cochrane reviews is yet to be inviolably established and the risk of missing eligible non-Cochrane reviews was too great for the potential benefits of time saved and greater efficiency of searching. It was therefore decided to take the comprehensive approach to identifying evidence, using the conventional search methods described below.

### 4.1.3. The need for an overview of diagnostic test of brief cognitive assessments for identifying dementia in primary care

The rationale behind this overview was to provide a clear and definitive answer to which brief cognitive assessment is most accurate for use in primary care as part of the toolkit for identifying dementia. Conducting an overview of diagnostic accuracy

systematic review evidence also presented the opportunity to explore and develop methods for carrying out overviews of diagnostic accuracy.

A number of Cochrane Reviews have explored the individual value of tests for dementia to general practitioners[198,199] to secondary care units such as memory clinics and acute care settings[200], and in the community[201-203] - or across a number of these settings[204-207]. These Cochrane reviews and other recent systematic reviews[99,120,208] have explored the diagnostic test accuracy of individual tests across a range of populations and settings. Given this current diversity of available systematic review evidence, researchers, policy-makers and practitioners would benefit from this evidence being summarised further for specific relevance and application to general practice. This specific and targeted summary of evidence is the focus of this overview.

## 4.2. Method

The aim of this overview was to summarise existing systematic review evidence for the diagnostic test accuracy of brief cognitive assessments, concentrating specifically on their use within a primary care setting (usually general practice). Primary care is the focus of many current discussions and policy-driven initiatives on increasing dementia diagnosis worldwide, yet the evidence underlying many of these debates is often unclear. A secondary aim is to summarise existing systematic review evidence for the diagnostic accuracy of informant-based brief cognitive assessments in the same setting.

### 4.2.1. Protocol

The overview protocol was published in the International Prospective Register for Systematic Reviews (PROSPERO) run by NIHR and the Centre of Reviews and Dissemination at University of York (registration reference CRD42015022078). The full protocol is shown in **Error! Reference source not found.**. The only change between protocol and review was in the search strategy. It was intended to search an evidence database being developed by the University of Exeter Medical School. The database however was not complete at the time of searching and therefore searches were restricted to the priority databases.

### 4.2.2. Inclusion and exclusion criteria

*Types of reviews*

Systematic reviews that examined the accuracy of brief cognitive assessments for identifying dementia in primary care, including relevant Cochrane reviews and protocols (in order to identify reviews in progess) published in The Cochrane Library were eligible for inclusion.

### 4.2.3. Types of participants

Studies that featured adults aged 18 years or over recruited from a primary care or general practice population were eligible for inclusion. Whilst patients who were selected on the basis of an existing diagnosis or condition which might reasonably be expected to feature in primary care (e.g. stroke) were not excluded, these patients were treated as a sub-group in the final analysis.

### 4.2.4. Target condition being diagnosed

All cause (non-differentiated) dementia was the target condition. Reviews that focused specifically on dementia subtypes such as Alzheimer's disease, vascular dementia and dementia with Lewy bodies were included, but reviews that focused on mild cognitive impairment (MCI) were excluded as this was not the condition of focus and both symptoms and aetiology are less well-defined than, and potentially significantly different to those of, dementia. Where reviews investigated both dementia and MCI, data referring to dementia were extracted and data referring solely to MCI were excluded.

### 4.2.5. Index tests

Index tests are the 'novel' test or tests introduced and compared against the established test or reference standard. All index tests within this overview are brief cognitive assessments. This term lacks accepted definition, but for the purposes of this overview 'brief' islimited to refer to tests which take up to 10 minutes to conduct. Analyses include the Mini Mental State Examination (MMSE); whilst the timings for this test are contested, it is one of the most commonly used tests and often characterised as taking 'up to' 10 minutes[64,209,210]. Brief cognitive assessments plus informant ratings were included where used, and a list of all included tests is presented within the results section.

#### 4.2.6. Reference standards

There is currently no gold standard test (i.e. close to 100% sensitivity and specificity) for identifying dementia in primary care. We therefore included all reference standard tests used within the relevant population and setting.

Reference standards are commonly selected on the basis of many variables such as common practice within individual clinics, practitioner preference, specialisation and experience of healthcare professionals and practice managers and are subject to changes in cost and fashion. Many of the globally-accepted reference standards such as the World Health Organisation-supported ICD and the DSM produced by the American Psychiatric Association are updated regularly; the DSM-5 (sometimes colloquially referred to as DSM-V) was released in 2013, and the ICD-11 is due for release by 2018.

### 4.3. Search methods for identification of reviews

#### 4.3.1. Electronic searches

The Cochrane Database of Systematic Reviews was searched using the search strategy shown in Appendix 1. Searches also covered EMBASE, MEDLINE and PsychInfo for eligible systematic reviews from inception until August 2015 and search strategies are also shown in Appendix 1. Hand searching, and forwards and backwards citation searches were carried out using Google Scholar for systematic reviews included at full text screening. Date and language restrictions were not applied, in line with advice from the Cochrane Handbook of Diagnostic Test Accuracy[211]. Where reviews had been updated, the latest available version was chosen. Additional papers were identified through Zetoc alerts and added in to the title/abstract screening. Two reviewers (HH and EK) independently screened titles/abstract and full texts in parallel and resolved discrepancies through discussion.

Updated searches were run on the Cochrane Database of Systematic Reviews (CDSR) in February 2016, and all eligible Cochrane Reviews of diagnostic accuracy are shown in **Error! Reference source not found.**. Updated searches were only run within CDSR as this was a simple and quick way to check whether previously-identified protocols may have been published as full reviews. As Cochrane Reviews are considered the highest quality of systematic review evidence available, this targeted and refined approach to update searches was considered rational within timescales.

### 4.3.2. Quality appraisal and risk of bias

As there is currently no clear guidance of the quality assessment for overview reviews of diagnostic test accuracy studies, and evidence of the best quality assessment tools is unclear, it was decided to carry out quality assessment using both the AMSTAR checklist[212] and the newer risk of bias assessment in systematic reviews (ROBIS) tool[213] to allow assessment of potential benefits and shortcomings of each measure in conducting an overview of diagnostic accuracy.

Review quality was assessed using the AMSTAR checklist[212] and the suitability of this tool to assess quality within an overview review of systematic reviews of diagnostic test accuracy was critically evaluated. The ROBIS tool[213] specifically assesses the degree to which review methods minimised risk of bias, and the degree to which the aims of the review were addressed by the review authors. As the ROBIS tool is newly developed, it was also assessed for its suitability for use in an overview of systematic reviews of diagnostic test accuracy.

The lead reviewer (HH) carried out the assessments, with the second reviewer (EK) checking assessments for consistency and agreement. Both tools were piloted independently by assessing two reviews[99,120] and comparing scoring decisions. When both reviewers were content with clear and rational approaches, the lead reviewer assessed the remaining reviews and the second reviewer spot-checked assessments. This was in order to maintain good systematic reviewing practice and consistency of approach whilst also making sure all substantive work was carried out by the lead reviewer.

### 4.3.3. Planned synthesis

Data were summarised in text and using summary tables. A summary of findings table was produced (Table 13**Error! Reference source not found.**) displaying review reference, component study reference, reference standard, index test, threshold, time taken, sample size and key demographic data, target condition, condition prevalence, accuracy data reported and notes. Additional summary tables were produced for comparisons at review level, non-DTA data reported, and individual brief cognitive assessments.

## 4.4. Overall results

It was not possible to determine which brief cognitive assessments were most accurate, or had the highest sensitivity and specificity at a given diagnostic threshold, when used as part of the process to identify dementia within a general practice population.

Clear conclusions could not be drawn across the included systematic reviews relating to the diagnostic accuracy of individual brief cognitive assessments. Broadly, this was due to: a lack of primary research; poor quality where the primary research did exist; and inconsistency in reporting results and primary data. Steps required to tackle these issues, and the costs of answering these questions weighed against their relative importance, are addressed in detail within the Discussion section.

### 4.4.1. Population

It was rare to find consistency amongst individual systematic review populations, as in many cases the population of interest was stated as ‘primary care’ or ‘general practice’ yet included study data from highly selected clinical populations. This is illustrated most clearly in Table 7 showing population data from included reviews against dementia prevalence where reported by individual study authors.

It is clear from Table 7 that whilst many systematic reviews stated their population of interest as ‘primary care’ or ‘general practice’, the dementia prevalence in included studies was as high as 51%[214]. This is close to what would be expected in the highly selected population of a memory clinic or residential home, rather than the far lower prevalence of around 6.5% seen in primary care[215]. Therefore with this level of variation and inconsistency within systematic reviews, there was little opportunity to statistically analyse results across included reviews in order to generate more meaningful comparisons.

Table 7 Population data from included reviews and prevalence (where reported) from studies included within those reviews

| Full systematic review reference | Verbatim stated review population | Full source study reference | Reported dementia prevalence |
|---|---|---|---|
| **Arevalo-Rodriguez, I., Smailagic, N., Roqué i Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., . . . Cullum, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). The Cochrane Library.** | Participants were recruited from: i) secondary care - outpatient clinic; ii) secondary care - memory clinics; & iii) populational sources (incl. primary care) | Meguro, K., Ishii, H., Kasuya, M., Akanuma, K., Meguro, M., Kasai, M.. Asada, T. (2007). Incidence of dementia and associated risk factors in Japan: The Osaki-Tajiri Project. Journal of the neurological sciences, 260(1), 175-182.<br>Modrego, P. J., Fayed, N., & Pina, M. A. (2005). Conversion from mild cognitive impairment to probable Alzheimer's disease predicted by brain magnetic resonance spectroscopy. American Journal of Psychiatry, 162(4), 667-675.<br>Modrego, P. J., & Gazulla, J. (2013). The predictive value of the memory impairment screen in patients with subjective memory complaints: a prospective study. The primary care companion for CNS disorders, 15(1).<br>Nakata, E., Kasai, M., Kasuya, M., Akanuma, K., Meguro, M., Ishii, H., Meguro, K. (2009). Combined memory and executive function tests can screen mild cognitive impairment and converters to dementia in a community: the Osaki-Tajiri project. Neuroepidemiology, 33(2), 103-110.<br>Xu, G., Meyer, J. S., Thornby, J., Chowdhury, M., & Quach, M. (2002). Screening for mild cognitive impairment (MCI) utilizing combined mini-mental-cognitive capacity examinations for identifying dementia prodromes. International journal of geriatric psychiatry, 17(11), 1027-1033. | - |
| **Brodaty, H., Low, L.-F., Gibson, L., & Burns, K. (2006). What is the best dementia screening instrument for general practitioners to use? The American Journal of Geriatric Psychiatry, 14(5), 391-400.** | Validated in "two distinct samples" or Inpatient or Outpatient settings All others validated in general practice, community or population settings | Solomon, P. R., & Pendlebury, W. W. (1998). Recognition of Alzheimer's disease: the 7 Minute Screen. Fam Med, 30(4), 265-271. | 50% |
| | | Kirby, M., Denihan, A., Bruce, I., Coakley, D., & Lawlor, B. A. (2001). The clock drawing test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly. International Journal of Geriatric Psychiatry, 16(10), 935-940 | 7% |
| | | Brodaty, H., Pond, D., Kemp, N. M., Luscombe, G., Harding, L., Berman, K., & Huppert, F. A. (2002). The GPCOG: a new screening test for dementia designed for general practice. Journal of the American Geriatrics Society, 50(3), 530-534. | 29% |
| | | 216 | 21% |
| **Carnero Pardo, C., Espejo Martínez, B., & Montoro Rios, M. (2009). Revisión sisteámtica y metaanálisis de la utilidad diagnóstica del** | Mix including Fase III Primary care | Carnero-Pardo C, Espejo-Martínez B, López-Alcalde S, Espinosa García M, Feria Vilar I, & L, M. N. (2008). ¿Es hora de jubilar al Mini-Mental? . Neurologia, 23, 648-649. | - |

| | | | |
|---|---|---|---|
| **Eurotest en la identificación de la demencia. Alzheimer(42), 14-22.** | | | |
| **Carnero-Pardo, C., Lopez-Alcalde, S., Allegri, R. F., & Russo, M. J. (2014). A systematic review and meta-analysis of the diagnostic accuracy of the Phototest for cognitive impairment and dementia. Dementia & Neuropsychologia, 8(2), 141-147.** | Primary care | Carnero-Pardo, C., Sáez-Zea, C., Montiel-Navarro, L., Feria-Vilar, I., & Gurpegui, M. (2011). Normative and reliability study of fototest. Neurología (English Edition), 26(1), 20-25. | 38% |
| **Creavin, S. T., Wisniewski, S., Noel-Storr, A. H., Trevelyan, C. M., Hampton, T., Rayment, D., . . . Milligan, R. (2016). Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. The Cochrane Library.** | Primary care | Brodaty, H., Pond, D., Kemp, N. M., Luscombe, G., Harding, L., Berman, K., & Huppert, F. A. (2002). The GPCOG: a new screening test for dementia designed for general practice. Journal of the American Geriatrics Society, 50(3), 530-534. | 29% |
| | | Lavery, L. L., Lu, S.-y., Chang, C.-C. H., Saxton, J., & Ganguli, M. (2007). Cognitive assessment of older primary care patients with and without memory complaints. Journal of general internal medicine, 22(7), 949-954. | 8.9% |
| | | Lourenço, R. A., & Veras, R. P. (2006). Mini-Mental State Examination: psychometric characteristics in elderly outpatients. Revista de Saúde Pública, 40(4), 712-719. | 25.7% |
| | | Pond, C. D., Mant, A., Kehoe, L., Hewitt, H., & Brodaty, H. (1994). General practitioner diagnosis of depression and dementia in the elderly: can academic detailing make a difference? Family Practice, 11(2), 141-147. | 15.5% |
| | | 217 | 21.4% |
| | | Cruz-Orduña, I., Bellón, J. M., Torrero, P., Aparicio, E., Sanz, A., Mula, N., . . . Olazarán, J. (2011). Detecting MCI and dementia in primary care: efficiency of the MMS, the FAQ and the IQCODE. Family Practice, 29(4), 401-406. | 9.4% |
| **Harrison, J. K., Fearon, P., Noel-Storr, A. H., McShane, R., Stott, D. J., & Quinn, T. J. (2014). Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general** | General practice/ primary care | Tokuhara, K. G., Valcour, V. G., Masaki, K. H., & Blanchette, P. L. (2006). Utility of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for dementia in a Japanese-American population. Hawaii Med J, 65(3), 72-75. | 7% |

| | | | |
|---|---|---|---|
| **practice (primary care) setting. The Cochrane Library.** | | | |
| **Lin, J. S., O'Connor, E., Rossom, R. C., Perdue, L. A., & Eckstrom, E. (2013). Screening for cognitive impairment in older adults: a systematic review for the US Preventive Services Task Force. Annals of Internal Medicine, 159(9), 601-612.** | Primary care | - (aggregated data) | |
| **Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. Journal of Psychiatric Research, 43(4), 411-431. doi:10.1016/j.jpsychires.2008.04.014** | Primary care | 218 | 6% |
| | | Borson, S., Scanlan, J. M., Chen, P., & Ganguli, M. (2003). The Mini-Cog as a screen for dementia: validation in a population-based sample. Journal of the American Geriatrics Society, 51(10), 1451-1454. | 6% |
| | | Brodaty, H., Pond, D., Kemp, N. M., Luscombe, G., Harding, L., Berman, K., & Huppert, F. A. (2002). The GPCOG: a new screening test for dementia designed for general practice. Journal of the American Geriatrics Society, 50(3), 530-534. | 29% |
| | | Grober, E., Hall, C., Lipton, R. B., & Teresi, J. A. (2008). Primary care screen for early dementia. Journal of the American Geriatrics Society, 56(2), 206-213. | 17% |
| | | Hooijer, C., Dinkgreve, M., Jonker, C., Lindeboom, J., & Kay, D. (1992). Short screening tests for dementia in the elderly population. I. A comparison between AMTS, MMSE, MSQ and SPMSQ. International Journal of Geriatric Psychiatry, 7(8), 559-571. | 4% |
| | | Kilada, S., Gamaldo, A., Grant, E. A., Moghekar, A., Morris, J. C., & O'brien, R. J. (2005). Brief screening tests for the diagnosis of dementia: comparison with the mini-mental state exam. Alzheimer Disease & Associated Disorders, 19(1), 8-16. | 42% |
| | | Kirby, M., Denihan, A., Bruce, I., Coakley, D., & Lawlor, B. A. (2001). The clock drawing test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly. International Journal of Geriatric Psychiatry, 16(10), 935-940. | 7% |
| | | O'Connor, D., Pollitt, P., Hyde, J., Fellows, J., Miller, N., Brook, C., & Reiss, B. (1989). The reliability and validity of the Mini-Mental State in a British community survey. Journal of psychiatric research, 23(1), 87-96. | 11% |
| | | 216 | 21% |

| | | | |
|---|---|---|---|
| | | Brayne, C., & Calloway, P. (1989). An epidemiological study of dementia in a rural population of elderly women. The British Journal of Psychiatry, 155(2), 214-219. | 8% |
| | | Clarke, M., Jagger, C., Anderson, J., Battcock, T., Kelly, F., & Stern, M. C. (1991). The prevalence of dementia in a total population: a comparison of two screening instruments. Age and ageing, 20(6), 396-403. | 51% |
| | | Cullen, B., Fahy, S., Cunningham, C. J., Coen, R. F., Bruce, I., Greene, E., . . . Lawlor, B. A. (2005). Screening for dementia in an Irish community sample using MMSE: a comparison of norm-adjusted versus fixed cut-points. International Journal of Geriatric Psychiatry, 20(4), 371-376. | 4% |
| **Mitchell, A. J., & Malladi, S. (2010). Screening and case finding tools for the detection of dementia. Part I: evidence-based meta-analysis of multidomain tests. The American Journal of Geriatric Psychiatry, 18(9), 759-782.** | Primary care | Brodaty, H., Pond, D., Kemp, N. M., Luscombe, G., Harding, L., Berman, K., & Huppert, F. A. (2002). The GPCOG: a new screening test for dementia designed for general practice. Journal of the American Geriatrics Society, 50(3), 530-534. | 28.5% |
| | | Brodaty, H., Kemp, N. M., & Low, L. F. (2004). Characteristics of the GPCOG, a screening tool for cognitive impairment. International Journal of Geriatric Psychiatry, 19(9), 870-874. | 28.5% |
| | | 146,153,219 | - |
| **Mitchell AJ, Malladi S. Screening and case-finding tools for the detection of dementia. Part II: evidence-based meta-analysis of single-domain tests. The American Journal of Geriatric Psychiatry. 2010;18(9):783-800.** | Primary care | Borson, S., Scanlan, J. M., Chen, P., & Ganguli, M. (2003). The Mini-Cog as a screen for dementia: validation in a population-based sample. Journal of the American Geriatrics Society, 51(10), 1451-1454.<br>Kirby, M., Denihan, A., Bruce, I., Coakley, D., & Lawlor, B. A. (2001). The clock drawing test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly. International Journal of Geriatric Psychiatry, 16(10), 935-940.<br>Heun, R., Papassotiropoulos, A., & Jennssen, F. (1998). The validity of psychometric instruments for detection of dementia in the elderly general population. International journal of geriatric psychiatry, 13(6), 368-380.<br>Kilada, S., Gamaldo, A., Grant, E. A., Moghekar, A., Morris, J. C., & O'brien, R. J. (2005). Brief screening tests for the diagnosis of dementia: comparison with the mini-mental state exam. Alzheimer Disease & Associated Disorders, 19(1), 8-16. | - |
| **Naqvi, R. M., Haider, S., Tomlinson, G., & Alibhai, S. (2015). Cognitive assessments in multicultural populations using the Rowland Universal Dementia Assessment Scale: a** | For culturally and linguistically diverse populations | Shaaban, J., Aziz, A. A., Abdullah, Z., & Ab Razak, A. (2013). Validation of the Malay Version of Rowland Universal Dementia Assessment Scale (MRUDAS) among Elderly Attending Primary Care Clinic. International Medical Journal, 20(5). | 20.4% |

| | | | |
|---|---|---|---|
| **systematic review and meta-analysis. Canadian Medical Association Journal, 187(5), E169-E175.** | | | |
| **Tsoi, K. K., Chan, J. Y., Hirai, H. W., Wong, S. Y., & Kwok, T. C. (2015). Cognitive tests to detect dementia: a systematic review and meta-analysis. JAMA internal medicine, 175(9), 1450-1458.** | Primary care | - (aggregated data) | |
| **Woodford, H., & George, J. (2007). Cognitive assessment in the elderly: a review of clinical methods. Qjm, 100(8), 469-484.** | General practice | Brodaty, H., Pond, D., Kemp, N. M., Luscombe, G., Harding, L., Berman, K., & Huppert, F. A. (2002). The GPCOG: a new screening test for dementia designed for general practice. Journal of the American Geriatrics Society, 50(3), 530-534. | 29% |

-, not reported; SR, systematic review; ref., reference; Incl., included; prev., prevalence

### 3.5.1. Accuracy

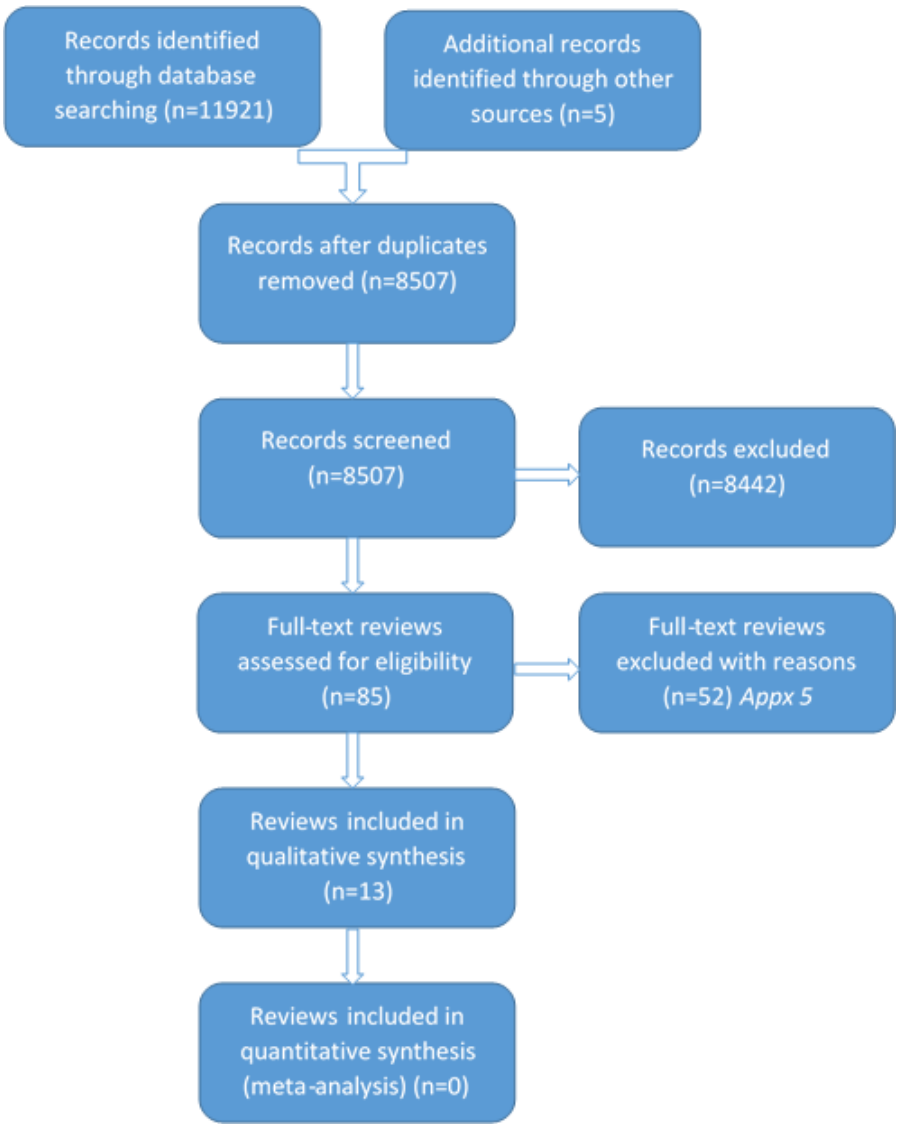
Where accuracy was clearly stated within systematic review evidence, reported accuracy differed between individual reviews. There was also little consistency in the reported diagnostic accuracy of the same brief cognitive assessments across systematic reviews. For the MMSE, one review[120] reported individual study sensitivity at unspecified thresholds ranging from 0.25 to 1.00, with specificity ranging from 0.54 to 1.00. The reporting of individual test accuracy data were inconsistent and non-existent in some cases[99,120]. This variation is illustrated most clearly in Table 12 **Error! Reference source not found.**, which summarises the evidence found across the eight included reviews which reported MMSE test accuracy data for identifying dementia within primary care. There was a great deal of disparity amongst the thresholds investigated, with one review apparently not reporting the MMSE threshold used at all[141] and one review reporting pooled data without reference to individual thresholds for MMSE – except to note that the most common threshold used within included studies (in 44% of cases) was <23.

This level of heterogeneity was found across all included studies, and made comparison across different reviews incredibly difficult, as even where there was consistency across populations, settings and tests, the thresholds used varied widely.

### 4.4.2. Detailed search results

The results of the searches are presented within a PRISMA flow diagram shown in **Error! Reference source not found.**. Full text exclusions are shown in **Error! Reference source not found.**. The characteristics of the 13 included systematic reviews are shown in the Summary of Findings table (Table 13**Error! Reference source not found.**) at the end of this chapter.

Figure 3 PRISMA flow diagram of overview



There was wide variation in fundamental factors of population, target condition, reference standards, index tests, thresholds and aspects not directly related to test accuracy such as test administration time. Results are discussed below across these factors to illustrate the degree of variation discovered, and the Summary of Findings table in Table 13**Error! Reference source not found.** provides a comprehensive overview where individual reviews can be compared and contrasted.

The 13 included systematic reviews included three Cochrane systematic reviews of diagnostic test accuracy[198,204,205] and ten non-Cochrane systematic reviews[99,120,125,141,148,220-222]. Excluding the two reviews[99,120] where individual study data was not reported, included evidence was derived from 28 separate studies. Fifteen studies featured in more than one of the included reviews (one study in four reviews[142], one study in three reviews[223], and four studies in two reviews[146,216,224,225]), illustrating a great deal of overlap between study-level evidence included within individual systematic reviews.

Most (8/13) of the assessments of diagnostic test accuracy focussed on the Mini Mental State Examination (MMSE) compared against a range of reference standards. There was considerable variation in assessed index test thresholds. The MMSE was assessed at ten thresholds from 17 to 27, with <23 (signifying that patients with an MMSE score below 23 were assessed as having cognitive impairment) the most common threshold used across all reviews. The most common reference standard used was DSM-IV, with a mixed range of reference standards used from simple clinical diagnosis to a combination of several tools. The majority of studies were concerned with the diagnosis of dementia reflecting our inclusion strategy, with one review focused specifically on conversion from MCI to dementia[204].

All thirteen included reviews were published since 2006, with 6 published since 2014[120,148,198,204,205,221]. Included studies were published between 1989 and 2013 (1989 to 1994, four studies; 1995 to 1999, three studies; 2000 to 2009, 14 studies; 2009 to 2013, five studies). Two included systematic reviews[99,120] only reported aggregated study data so it was not possible to identify individual studies from these two systematic reviews. This individual study level data is missing from the rest of the results and analysis. Where reported, median total study sample size ranged from 49 participants[148,226] to 1178 participants[218,222].

Of the component studies: nine were based in the USA[145,218,224,227-230] (with two studies [219,225] featured in two separate reviews[222,231]); six in Spain[153,217,232-235]; three in Australia (with one study[142] referenced in five separate reviews[125,141,205,222,236]); two in Japan[237,238]; two in the United Kingdom[239,240]; two in the Netherlands[146,241] (with one study [146] referenced in two reviews[141,222] and one study[241] referenced in another two separate reviews[222,236]); one in Grenada[242]; one in Ireland (with one study[223] referenced in three separate reviews[222,231,236]); one in Malaysia[226]; one in Brazil[243]; and one study based in Germany[244]. Two of the included systematic reviews[99,120] only reported pooled study data, so population data for individual studies were not available.

The low numbers of UK-based population data at the study level is a limitation worth noting for this overview, where the focus is on general practice populations relevant to the UK primary care setting. This limits the generaliseability of findings yet further, and is important to note for the cautious application of current evidence to a UK or mainland European general practice population.

### 4.4.3. Target condition

Dementia, all cause dementia, early dementia and dementia in normal and depressed elderly were the target conditions within ten reviews[99,125,141,148,198,205,221,222,245]. One review[141] concentrated on mixed severity dementia, dementia in multicultural populations and dementia in people with a low educational level. In one review[204], conversion rates were the focus from mild cognitive impairment (MCI) to dementia and Alzheimer's disease or probable Alzheimer's disease, with one component study[227] addressing conversion from subjective memory complaints to Alzheimer's disease, vascular dementia, dementia with Lewy bodies or frontotemporal dementia. Alzheimer's disease, dementia in normal and depressed elderly and dementia were the focus of one review[236]. Two reviews[99,120] reported aggregate study data from a number of different countries and settings incorporating community, senior communities, assisted living facilities, clinic (undefined), hospital, primary care, and 'other' (undefined). One of these reviews[99] reported a mean participant age range of 65 to 91 years, and dementia prevalence on a range of 1.2% to 47.1%.

### 4.4.4. Reference standards

*Suitability of reference standards*

Whilst our inclusion criteria purposefully included all-cause dementia, reference standards of included systematic reviews included highly specific tools designed to identify dementia subtypes such as the NINCDS-ADRDA criteria for Alzheimer's disease, NINCDS-AIREN for Vascular dementia, the McKeith criteria for dementia with Lewy bodies and the Lund and Manchester criteria for frontotemporal dementia.

As subtypes are generally not identified at the level of primary care, this provided an early indication that included data (at least) at the study level had included specific clinical populations not necessarily suited to a general practice population raises an early query around the clinical suitability of studies assessed within ostensibly community or primary care-focussed reviews.

*Interconnected brief cognitive assessments*

Some tools used as reference standards are interrelated and use elements of other measures within their structure. For example, the Cambridge Mental Disorders of the Elderly Examination (CAMDEX) has three main sections: A structured clinical patient interview to assess their present state, past history and family history; a battery of cognitive tests; and a structured interview with a relative or other informant to gain information on the patient[246]. The Cambridge Cognitive Examination (CAMCOG) is a self-contained measure consisting of the battery of cognitive tests taken directly from the CAMDEX[247]. The CAMCOG (and thus CAMDEX) incorporates a section with all nineteen questions from the Mini Mental State Examination (MMSE)[248]. One review[236] includes study data that used CAMCOG as an index test with CAMDEX as a reference standard in one study[142], and MMSE as an index test with CAMDEX as reference standard in another study. This clearly raises questions of incorporation bias when there is overlap and duplication between the reference standard and index tests.

Similarly, the Short and Sweet Screening Instrument (SASSI or SAS-SI) used as a reference standard within a study[218] included in the 2009 systematic review by Mitchell[222] consists of the MMSE, category fluency for animals and a temporal orientation test[248]. As the index test within this study was the MMSE, here again is clear potential for the introduction of systematic bias.

The Mini-Cog combines the Clock Drawing Test (CDT) and the three-item word memory task[224], and although it was not used as a reference standard in any of the

included studies, diagnostic performance was compared against the CDT and three-word recall task within one of the component studies[224].

*Reference standards within included reviews*

Five reviews[141,205,222,236] incorporated a range of reference standards from:

- clinical diagnosis using NINDCS-ADRDA[228],
- clinical diagnosis using CAMDEX and DSM-IV[142]
- GP diagnosis using CAMDEX and GMS-AGECAT[216],
- DSM-IV plus 'expert consensus'[230],
- to using Mini-Cog with DSM-III-R and NINCDR-CERAD[224],
- GMS-AGECAT plus CAMDEX[146],
- DSM-III and ICD-10[249],
- DSM-III-R with GMS-AGECAT[223],
- DSM-III-R and NINCDR-CERAD[244],
- DSM-IV and ICD-10[243],
- DSM-III-R and IPA/WHO criteria[153],
- SASSI[218],
- GMS-AGECAT[216],
- AGECAT dementia[250],
- DMS-III-R[219,225],
- DSM-IV[142,251],
- DSM-IV-R[232,252,253],
- CAMDEX[214,239,240] or
- CDR[229]

Three reviews[125,148,245] solely reported DSM-IV as a reference standard (from studies[142,217,226] with one review[125] also using CAMDEX[142]).

One review[204] primarily incorporated the NINCDS-ADRDA criteria as reference standard for Alzheimer's disease, with two of the component studies also using NINCDS-AIREN as a reference standard for Vascular dementia, the McKeith criteria for dementia with Lewy bodies and the Lund and Manchester criteria for frontotemporal dementia[227] and NINCDS-ADRDA for probable Alzheimer's disease, NINDS-AIREN for possible Alzheimer's disease with cardiovascular disease, NINDS-AIREN for probable Vascular dementia, consensus guidelines for dementia with Lewy bodies and the Lund & Manchester Groups criteria for frontotemporal dementia[238]. One study[238] also used DSM-IV and CDR 1+ as a reference standard for conversion from MCI to dementia.

One review[198] described a reference standard of clinical dementia diagnosis using Benson and Cummings criteria, informed by cognitive testing (CASI, MMSE and Clock-Drawing Task), interview data and an assessment of function[145].

Two reviews reported aggregated data across all included studies[99,120]. In one of these two reviews[120] the reference standards used in primary care settings could not be disaggregated. Authors of the other review[99] selectively reported the most common reference standards as criteria from the Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III), the DSM-IV, or the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association. Formal diagnosis was based on a combination of history, examination, neuropsychological testing, and expert consensus. One review[221] reported "clinical diagnosis" as the reference standard without reference to any tool[242].

Index tests

The MMSE was most frequently assessed, appearing in 8 reviews[99,120,125,141,204,205,222,236], with Eurotest, Fototest, BIMCT, John Brown test, Verbal Fluency, Orientation, RUDAS and PCL all featuring in only one review each (see Table 8**Error! Reference source not found.**). The information within this table is not an unbiased reflection of frequency, as two of the reviews[99,120] only report aggregated study data - meaning the studies which had reported data on individual index tests could not be identified.

Table 8 Frequency of index test use

| Index test | Number of review inclusions | Number of unique study inclusions |
|---|---|---|
| MMSE | 8[99,120,125,141,204,205,222,236] | 22[142,218,224,240x3 146,214,216,223,225,227,229,230,232,234,235,238,239,243,249,250,252] |
| GPCOG | 5[99,120,125,141,236] | 2[142,251] |
| CDT | 4[99,120,141,236] | 2[223,224] |
| Mini-Cog | 4[99,120,125,236] | 3[144,224,254-257] |
| IQCODE | 3[99,120,198] | 1[145] |
| MIS | 2[22,23] | 5[149,256,258-260] |
| SPMSQ | 2[99,141] | 2[146,219] |
| 7MS | 2[99,236] | 1[228] |
| 3 word recall | 1[141] | 2[224,244] |
| AMTS | 1[141] | 2[142,146] |
| Eurotest | 1[220] | 1[217] |
| Fototest | 1[221] | 1[242] |
| BIMCT | 1[141] | 1[219] |
| John Brown test | 1[141] | 1[225] |
| Verbal Fluency | 1[141] | 1[225] |
| Orientation | 1[141] | 1[225] |
| RUDAS | 1[148] | 1[226] |
| PCL | 1[141] | 1[153] |

MMSE, mini mental state examination [standard, Spanish and Portuguese editions]; IQCODE, Informant Questionnaire on Cognitive Decline in the Elderly; GPCOG, General Practitioner assessment of Cognition; CDT, clock drawing test; MIS, memory impairment screen; SPMSQ, Short portable mental status questionnaire; AMTS, Abbreviated Mental Test Score; 7MS, 7 minute screen; BIMCT, Blessed Information Memory Concentration Test; RUDAS, The Rowland Universal Dementia Assessment Scale; PCL, Prueba cognitive de leganes.

Three reviews[204,205,222] solely assessed the performance of the MMSE at various thresholds. A range of index tests were assessed as indirect comparisons in three reviews[141,222,236], with index tests including the 7 minute screen[228], CDT[223], 3 word recall[224,244], GPCOG[142], GPCOG 'combined' *[N.B. combination not specified]* [251], PCL[153], SPMSQ and Blessed Information Memory Concentration Test[219], SPMSQ, MSQ, AMTS and AMTS combined with MSQ[146], John Brown test, verbal fluency – animals, orientation[225], and MMSE[216]. One review[220] assessed performance of the Fototest[217], and one related review[221] assessed performance of the Eurotest[242].

IQCODE was assessed at various thresholds between 3.2 and 3.7 in one review[200]. The RUDAS tool was assessed[148] in one review with a 23 out of 30 cut-off, meaning patients had to score below 23 to be assessed as having dementia. GPCOG, AMT and MMSE were assessed as direct comparisons in one review[125]. One review[120] made indirect comparisons between MMSE, CDT-Shulman, CDT-Sunderland, MiniCog, MIS, verbal fluency, AMT, GPCOG, MoCA, ACE-R, IQCODE (short and long

versions) and 3MS lthough as data were aggregated it was unclear which if any of these index tests assessed overlapping populations. Finally, CDT, MiniCog, MIS/MIS-T, MSQ/SPMSQ, verbal fluency, MMSE, AMT, FCSRT, 7MS, TICS and IQCODE were all assessed indirectly in one review[99]. As in the previous review the data was aggregated across studies and it was not possible to separate between different studies and populations, so there may be duplucation across participants within these data.

### 4.4.5. Other test features recorded in included systematic reviews

Many of the included systematic reviews noted factors beyond test accuracy in varying degrees of detail, and these were recorded for this overview at the stage of data extraction. These items were recorded within systematic reviews directly as follows: time taken to administer the test[99,105,120,125,141,148,221,231,236]; language/culture bias[125,148,236,261]; education bias[125,148,236]; face validity[236,261]; internal consistency[148,236]; interrater reliability[148,236]; test-retest reliability[148,236]; practicability for general practice[99,125]; Clinician preference[148]; and immigration status[148]. These aspects are noted here for completeness and addressed in detail in the Discussion section. The usefulness of diagnostic test accuracy as a measure of a test's value, and factors of value beyond test accuracy across all brief cognitive tests, are also addressed within the Discussion.

Table 9 What did the reviews say about test accuracy of the MMSE?

| Ref. | Reference standard | MMSE thresholds | Setting | Main findings | Other outcomes considered beyond test accuracy |
|---|---|---|---|---|---|
| 204 | NINCDS-ADRDA or DSM or ICD criteria for Alzheimer's disease dementia; McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; and NINDS-AIREN criteria for vascular dementia | ≤ 21,<br>≤ 26,<br>≤ 28,<br>≤ 29 | Participants were recruited from:<br>i) secondary care - outpatient clinic (n = 3);<br>ii) secondary care - memory clinics (n = 6) and<br>iii) populational sources (n = 2) | Only dealt with conversion from MCI to dementia.<br>No evidence supporting a substantial role of MMSE as a stand-alone single-administration test in the identification of MCI patients who could develop dementia.<br>Clinicians could prefer to request additional and extensive tests to be sure about the management of these patients. An important aspect to assess in future updates is if conversion to dementia from MCI stages could be predicted better by MMSE changes over time instead of single measurements. It is also important to assess if a set of tests, rather than an isolated one, may be more successful in predicting conversion from MCI to dementia. | - |
| 236 | Clinical diagnosis/clinical diagnosis combined with e.g. DSM-IV | ≤23 | Validated in "two distinct samples" or Inpatient or Outpatient settings<br>All others validated in general practice, community or population settings | Sensitivity = 69 (95% CI, 0.66-0.73)<br>Specificity = 89 (95% CI, 0.87-0.92)<br>PPV = 0.63 (95% CI, 0.58 – 0.67)<br>NPV = 0.92 (95% CI, 0.90 – 0.94)<br>Misclassification = 15% (calculated using DAGStag programme)<br>The authors recommend that GPs consider using the GPCOG, Mini-Cog or MIS when screening for cognitive impairment or for case detection. Quick and easy to administer whilst having psychometric properties similar to the MMSE. | Time taken<br>Face validity<br>Internal consistency<br>Education Bias<br>Language/Culture Bias<br>Interrater Reliability<br>Test-Retest Reliability<br>Ease of administration<br>Practicability for general practice |

| | | | | | |
|---|---|---|---|---|---|
| 205 | Clinical diagnosis defined by DSM, ICD or CRS. | ≤17<br>≤18<br>≤19<br>≤20<br>≤21<br>≤22<br>≤23<br>≤24<br>≤25<br>≤26 | Primary care | The authors could not estimate summary DTA in primary care due to insufficient data.<br>"The MMSE contributes to a diagnosis of dementia in low prevalence settings, but should not be used in isolation to confirm or exclude disease. We recommend that future work evaluates the diagnostic accuracy of tests in the context of the diagnostic pathway experienced by the patient and that investigators report how undergoing the MMSE changes patient-relevant outcomes", e.g. time to diagnosis, initiation of treatment/care package, additional testing and place of care. | - |
| 222 | Clinical diagnosis using DSM III, IIIR or IV | ≤21<br>≤22<br>≤23<br>≤24<br>≤25<br>≤26<br>≤28 | Primary care | In those studies conducted purely in primary care:<br>Sensitivity = 0.78<br>Specificity = 0.88<br>PPV = 0.54<br>NPV = 0.96<br>In non-specialist settings, the MMSE was best at ruling out dementia, achieving about 29/30 correct reassurances with less than three false negatives out of every 100 screens. MMSE offers modest accuracy with best value for ruling-out a diagnosis of dementia in community and primary care. For all other used it should be combined with or replaced by other methods. | - |
| 141 | Clinical diagnosis using DSM/IPA-WHO criteria/GMS-AGECAT 4 or 5+ CAMDEX | NR | Primary care | The authors conclude that the AMTS was preferable to the MMSE for case-finding, whereas the SMPSQ was inferior.<br>For screening, the MMSE was optimal and…the best tool for primary care physicians who want a rule in and rule out tool, if length is not a major consideration.<br>"at least 30 well-studied alternatives to the MMSE exist and several seem to be briefer but no less accurate than the MMSE…In primary care…these | Comparative accuracy conducted with 5 instruments (AMTS/MSQ, MSQ, WINDSET, PCL and AMTS). Time taken given in all cases. |

| | | | | | |
|---|---|---|---|---|---|
| | | | | methods would help detect on 13 of 20 possible cases but rule out 18 of 20 healthy individuals (overall correct in 17 out of 20) suggesting these brief batteries may be ideally suited to a first step screen followed by a more comprehensive in those who screen positive" (p. 779). | |
| 99 | DSM criteria and NINCDS-ADRDA | Various | Primary care and community (primary care-relevant populations) | Review authors reported on MMSE (n = 12 348). …"The best-studied instrument was the MMSE. Pooled estimates across 14 studies (n = 10 185) resulted in sensitivity of 88.3% (95% CI, 81.3% to 92.9%) and specificity of 86.2% (CI, 81.8% to 89.7%) for the most commonly reported cut points of 23/24 or 24/25.<br>Review authors' conclusions:<br>"…only a handful of instruments have been studied in more than 1 study applicable to primary care. Although the MMSE is the best-studied instrument, it has the longest administration time and is not available for public use without cost. | Other assessments of benefits and harms, caregiver interventions, non-pharmacological interventions aimed at the patients |
| 120 | DSM, ICD, NINCDS-ADRDA, CAMDEX, GMS-AGECAT, CERAD, CDR (not reported individually) | NR | Primary care | From the authors –<br>"With different cut-off threshold values, we found considerable variation in the sensitivity and specificity estimates reported by individual studies. The sensitivities ranged from 0.25 to 1.00, and the specificities ranged from 0.54 to 1.00. The heterogeneity among studies was large, with I2 statistics for sensitivity and specificity of 92% and 94%, respectively…The combined data in the bivariate random-effects model gave a summary point with 0.81 sensitivity (95% CI, 0.78-0.84) and 0.89 specificity (95% CI, 0.87-0.91). The HSROC curve was plotted with a diagnostic odds ratio of 35.4, and the AUC was 92% (95%CI,90%-94%)" (p.E4). | Administration time, number of questions, components of screening tests |

| | | | | | |
|---|---|---|---|---|---|
| 125 | CAMDEX, DSM-IV | <8<br><25 | General practice | Review authors concluded that the GPCOG outperformed the MMSE in a general practice sample. “However, the composition of the test was changed post hoc, and the available data are incomplete”[my bold/italics]<br>“…the limited cognitive domains that all these tests assess may make them prone to miss non-Alzheimer’s dementias” (p.475) | Other assessment domains (language, attention, memory, visuospatial skills, executive function) and monitoring disease progression discussed. |

CAMDEX, Cambridge Mental Disorders of the Elderly Examination; CERAD, Consortium to Establish a Registry for Alzheimer’s Disease; CDR, Clinical Dementia Rating; DSM, DSM-IV, MMSE, mini mental state examination; GPCOG, The General Practitioner assessment of Cognition; CI, confidence interval; ICD, International Classification of Diseases; NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer’s Disease and Related Disorders Association; GMS-AGECAT, Geriatric Mental State – Automated Geriatric Examination for Computer Assisted Taxonomy; n, number in sample; AMTS, Abbreviated Mental Test Score; MSQ, Mental Status Questionnaire; PCL, Prueba cognitive de leganes *[Leganés cognitive test]*; IPA-WHO, International Psychogeriatric Association World Health Organisation criteria; PPV, positive predictive value; NPV, negative predictive value; ICD, CRS, MCI, mild cognitive impairment; NINDS-AIREN, National Institute of Neurological Disorders and Stroke and Association Internationales pour la Recherche et l´ Ensignement en Neurosciences

### 4.4.6. Quality assessment across all reviews

Table 10**Error! Reference source not found.** shows all quality assessment scores using the AMSTAR tool, and the full tool is shown in **Error! Reference source not found.**. Across the reviews, Cochrane reviews[198,204,205] were of a higher methodological quality (all scored 10 out of 11) whereas the non-Cochrane reviews[120,125,141,148,190,221,222,236,245] were of moderate or low quality (mean score 4.8 out of 11, range 0-7).

All three Cochrane reviews prespecified the research question and inclusion criteria in the form of a published protocol, and gave clear evidence of duplicate study selection and data extraction with a comprehensive literature search including grey literature searches. All these provided a clear list of included and excluded studies, containing characteristics of included studies and quality assessments reported and incorporated into review conclusions. It was judged that the methods used for combining findings were appropriate in all cases, and conflicts of interests featured in each Cochrane review. The likelihood of publication bias using graphical aids and/or statistical methods was not assessed in any of the included Cochrane reviews. However, as all the included Cochrane reviews were systematic reviews of diagnostic test accuracy and the nature of publication bias in test accuracy studies is still unclear, this may be a less applicable measure within this methodological context compared to systematic reviews of interventions. This appears to be a weakness of existing tools and worth amending in further revisions if they can be tailored for test accuracy reviews.

None of the non-Cochrane reviews gave evidence of an ‘a priori’ design with prespecificed research questions or inclusion criteria. It was not possible to assess whether there was duplicate study selection and data extraction in five out of ten non-Cochrane reviews as there were insufficient data reported. Authors were not contacted as the lack of data reported across a number of different areas indicated systemic issues with reporting quality rather than individual authorship decisions. Comprehensive literature searches were performed in eleven out of thirteen reviews, with two reviews[125,236] assessed as low quality for the data presented within the reviews on their literature searches.

Four reviews clearly reported included/excluded studies[99,198,204,205]. Conflicts of interest reporting was assessed as poor in seven reviews[120,125,141,148,222,236].

### 4.4.7. Risk of bias

The ROBIS tool[213] was used to assess the risk of bias within included systematic reviews. This recently-developed tool is adapted for different systematic review methodologies, and incorporates questions tailored for systematic reviews of diagnostic test accuracy. Table 11**Error! Reference source not found.** shows all the ROBIS scores for included systematic reviews, and the complete tool is available in **Error! Reference source not found.**.

Overall, six reviews[99,141,148,198,204,205] were judged to be at low risk of bias, with one review[141] at medium risk of bias and six[120,125,221,222,236,245] reviews assessed at high risk of bias overall.

Three reviews[198,204,205] were judged at low risk of bias across all four domains:

1. Study eligibility criteria
2. Identification and selection of studies
3. Data collection and study appraisal
4. Synthesis and findings.

All three of these reviews[198,204,205] were Cochrane reviews, and it should be noted that the ROBIS tool was developed in line with the Cochrane reporting guidelines contained within the Cochrane Handbook for Systematic Reviews of Interventions[262] and Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[263]. It is therefore unsurprising that there is significant resonance between systematic reviews following the prescriptive Cochrane method, and a risk of bias assessment tool designed using the same Cochrane approach.

Two reviews[120,236] were assessed as high risk of bias across all four of these domains, with particular concerns around lack of information on predefined objectives, data collection and no clear accounting for heterogeneous results[236].

One review[120] pooled sensitivity and specificity data without giving a rationale, and accuracy was combined without reference to appropriate populations, settings and administration times. Most curiously the authors of this review generated a hierarchical summary of Receiver-operating Characteristics (HSROC) graphic, but *removed* the confidence region, thus giving a misleading view of an inappropriate combination of data.

Table 10. AMSTAR quality assessment scores across all reviews

| AMSTAR question | Harrison 2014 | Aravelo-Rodriguez 2015 | Creavin 2016 | Lin 2013 | Carnero-Pardo 2014 | Carnero-Pardo 2009 | Mitchell 2010a | Mitchell 2010b | Naqvi 2015 | Tsoi 2015 | Mitchell 2009 | Woodford 2007 | Brodaty 2006 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Was an 'a priori' design provided? | Yes | Yes | Yes | No | CA | No | No | No | No | No | No | No | No |
| 2. Was there duplicate study selection and data extraction? | Yes | Yes | Yes | Yes | CA | CA | No | No | Yes | Yes | CA | CA | CA |
| 3. Was a comprehensive literature search performed? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | No |
| 4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? | Yes | Yes | Yes | CA | Yes | Yes | No | No | CA | No | No | No | No |
| 5. Was a list of studies (included and excluded) provided? | Yes | Yes | Yes | Yes | No | No | No | No | No | No | No | No | No |
| 6. Were the characteristics of the included studies provided? | Yes | Yes | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes | No |
| 7. Was the scientific quality of the included studies assessed and documented? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No |
| 8. Was the scientific quality of the included studies used appropriately in formulating conclusions? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | No | No | No |
| 9. Were the methods used to combine the findings of studies appropriate? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | No | N/a |
| 10. Was the likelihood of publication bias assessed? | N/a | No | No | Yes | No | No | Yes | Yes | No | No | No | No | No |
| 11. Was the conflict of interest included? | Yes | Yes | Yes | No | Yes | Yes | No | No | No | No | No | No | No |
| AMSTAR overall ratings | **10/11** | **10/11** | **10/11** | **7/11** | **6/11** | **6/11** | **6/11** | **6/11** | **6/11** | **4/11** | **4/11** | **3/11** | **0/11** |

CA, cannot assess (insufficient data); N/a, not applicable

Table 11. ROBIS risk of bias assessment across all included reviews

| | Creavin 2016 **[C]** | Harrison 2014 **[C]** | Aravelo-Rodriguez 2015 **[C]** | Lin 2013 | Mitchell 2010a | Naqvi 2015 | Mitchell 2010b | Mitchell 2009 | Woodford 2007 | Carnero-Pardo 2014 | Carnero-Pardo 2009 | Brodaty 2006 | Tsoi 2015 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Overall risk of bias in the review** | ☺ | ☺ | ☺ | ☺ | ☺ | ☺ | 😐 | ☹ | ☹ | ☹ | ☹ | ☹ | ☹ |
| 1. Study eligibility criteria | ☺ | ☺ | ☺ | ☺ | ☺ | ☹ | ☹ | 😐 | 😐 | ☹ | ☹ | ☹ | ☹ |
| 2. Identification and selection of studies | ☺ | ☺ | ☺ | ☺ | ☺ | ☹ | ☹ | 😐 | 😐 | ☺ | ☹ | ☹ | ☹ |
| 3. Data collection and study appraisal | ☺ | ☺ | ☺ | 😐 | 😐 | ☺ | 😐 | ☹ | ☹ | ☹ | ☹ | ☹ | ☹ |
| 4. Synthesis and findings | ☺ | ☺ | ☺ | ☺ | ☺ | ☺ | 😐 | ☹ | ☹ | ☹ | 😐 | ☹ | ☹ |

| | |
|---|---|
| ☺ | Low risk of bias rating |
| 😐 | Medium risk of bias rating |
| ☹ | High risk of bias rating |
| **[C]** | Cochrane review |

## 4.5. Discussion and relevance

### 4.5.1. The evidence is not clear

Despite being conducted to a high standard, with a PROSPERO-registered protocol published open access in a peer-reviewed journal, designed and conducted in line with best evidence recommendations and drawing on an array of topic and methodological expertise, this overview found no evidence for one brief cognitive assessment which demonstrates clear evidence of superior diagnostic accuracy when used as part of the process for identifying dementia in general practice.

As the main aim of this overview was to address this single, seemingly-simple question, this was the most surprising and disappointing finding of this synthesis of existing evidence. To be clear, the finding was not a simple lack of evidence – which would account for continued uncertainty and be clearly addressed by a direct call for further research to address this specific issue.

There was a great deal of evidence identified within this overview (as demonstrated in the Summary of findings at the systematic review level shown in Table 13**Error! Reference source not found.**), but data suffered from three main issues:

Firstly, there is insufficient evidence at the primary study level which specifically addresses cognitive assessments designed and validated for use in a general practice population. Nor does sufficient evidence take account of standard general practice setting parameters, e.g. consultation length, comorbidities, mixed or indirect presentation, patient age, lack of clinical specialty, potential referral by a relative and presence of an informant. Much of the existing evidence claims to refer to such limits, but does not present clear empirical evidence for them.

The second issue is that poorly-conducted primary studies are ill-suited to answer the questions they purport to address. Of the systematic reviews included within this overview, the highest quality reviews[105,198,204] were unable to provide clear, unambiguous guidance on which assessments were most accurate for diagnosing dementia within a general practice population. This was despite identifying evidence eligible for inclusion within the included systematic reviews, although concerns were raised across the included systematic reviews on issues such as low study numbers[198,204], lack of direct comparisons between cognitive assessments[236], variation across study populations[120,125,221], and concerns around patient selection[105]

and generaliseability[99,148]. Within the overview, all included systematic reviews consisted of studies adopting various designs. Systematic reviews were part of the overview inclusion criteria, but in the absence of specific guidance on conducting overviews of diagnostic test accuracy the study designs of all the component studies were included within the overview. This is a key point for consideration by those creating guidelines for the conduct of overviews.

Finally, there is a problem with the quality of evidence at the level of systematic review, which does not clearly reflect the primary research. As Table 7**Error! Reference source not found.** illustrates, whilst included systematic reviews specified population inclusion criteria relating to general practice or primary care, dementia prevalence reported within primary studies of one review[222] ranged from 4%[250] to 51%[214] – far from UK general practice prevalence rates of around 6%[13]. Some reviews made recommendations on individual tools, but these recommendations were based upon indirect study comparisons, or studies where reported data were pooled without reporting disaggregated factors such as population data. In effect, study data from memory clinic and hospital populations were combined with study data from general practice populations, and this resulted in findings that could not be generalised to our population of interest. Whilst the inclusion criteria for this overview required that the population of interest (general practice or primary care) was explicitly stated within the review, this relied upon included reviews containing eligible studies.

### 4.5.2. What to do with current evidence

Aside from the issues raised above, there are clear messages to take from this overview of the current evidence. The accuracy of the five brief cognitive assessments most frequently featured within the included reviews within this overview (MMSE, GPCOG, CDT, Mini-Cog, IQCODE) are summarised below.

*MMSE accuracy*

Within this up-to-date overview with all reviews conducted since 2006, the most commonly-used brief cognitive assessment within research studies was the MMSE, appearing in eight of the included systematic reviews and twenty-two of the unique component studies (review and study references shown in Table 12**Error! Reference source not found.**). Despite its continuing ubiquity as a research tool, the MMSE is not suitable for use in clinical practice and should be avoided within general practice simply due to the length of time it takes to administer (7.3±5 minutes[144], longer than

practicable for the standard mean general practice appointment length of 10.7 minutes (6.7 standard deviation)[132]). In terms of diagnostic performance, the value of MMSE as a cognitive assessment tool for use in general practice is limited[241]. The MMSE has been found to overestimate impairment in people over 60 years old, and in those with lower levels of education[264]. Across included systematic review evidence and beyond[265,266], high levels of heterogeneity in MMSE accuracy makes clinical practice recommendations highly challenging.

Within this overview, the accuracy of MMSE in detecting dementia at different thresholds was highly varied, with a general pattern of higher sensitivity and lower specificity observed across reviews using lower thresholds of 19 to 22, and a tipping point of lower sensitivity and higher specificity emerging more strongly as thresholds increased from 23 to 27. However, this was inconsistent across all results and incorporated a very large range of dementia prevalence, ranging from 4% to 51% within a single systematic review[222].

*GPCOG accuracy*

The second most commonly featured brief cognitive assessment within the overview was the GPCOG, assessed in five[99,120,125,141,236] out of the thirteen included systematic reviews and two[142,251] component studies (one study[142] featured in four separate reviews). As one of the newer brief cognitive assessment tools, the GPCOG is less established in clinical and research applications but is growing in popularity in practice guidelines[266] (see Chapter 3). The GPCOG is often presented within research literature as designed specifically for use in general practice and a suitable tool due to clinical acceptability and limited bias for education, ethnicity and gender[130,248] with similar diagnostic performance to the MMSE[236,251]. Evidence identified in this overview did not refute this claim, but there was very limited evidence of direct comparisons between GPCOG and MMSE and indirect comparison data should be viewed with caution due to the likely variation introduced when viewing two different study populations, settings and testing protocols.

One review exploring the best dementia screening instrument for general practitioners [236] reported GPCOG two step sensitivity as 0.85 (95% CI 0.76-0.92) and specificity as 0.86 (95% CI 0.81-0.91). Another review of multi-domain tests for the detection of dementia assessed the GPCOG from two studies[142,251] but did not report sensitivity or specificity from these studies. Two other reviews[99,120] reported pooled accuracy data

across studies[142] with one recurrent GPCOG reference referring to another systematic review[236] (included both within this overview and within the systematic review of direct comparisons). One of these reviews reported pooled data for GPCOG, with pooled sensitivity of 0.92 (95% CI 0.81-0.97) and pooled specificity of 0.87 (95% CI 0.83-0.90).

Direct comparisons between tools were only reported in one review of clinical methods for cognitive assessment in the elderly[125] which reported GPCOG sensitivity of 0.85 and specificity of 0.86. This aligns with reporting of the study[142] included in the three other reviews described above[99,141,236].

*CDT accuracy*

The Clock Drawing Test (CDT) featured as an index test in four of the included systematic reviews[99,120,231,236], with one component study[223] featuring in three reviews[99,231,236]. Two reviews[99,120] reporting aggregated data from a pool of non-differentiated studies (one review[99] referencing seven studies[223,229,257,258,267-269]; one review[120] simply referenced the original verification studies for CDT[270,271] and noted that they included data “from nine studies” for each verification study).

There was no evidence found of direct comparisons between the CDT and other tools. In the two reviews that reported pooled data[99,120] it was not possible to disaggregate study level evidence to see if direct comparison data were available.

The two reviews[231,236] which include evidence from the same study [223] are consistent in reporting that the study took place in a general practice population of normal and depressed elderly people, with an average age of 77 years old[236] and a dementia prevalence of 7%. One review reported a sample population of 564 people[236] whereas the other review reported a population of 648 people, 41 with dementia[231]. The reference standard used was clinical diagnosis combined with DSM-IV criteria. Both reviews reported CDT sensitivity from the single study of 0.76 (95% CI 0.60-0.88) and specificity of 0.81 (95% CI 0.77-0.84).

One review[231] also reported CDT accuracy data from a study[224] conducted in community dwelling adults aged 65 years and over. CDT was compared to the reference standard of a field-modified version of the Consortium to Establish a Registry for Alzheimer's disease (CERAD) protocol and the Clinical Dementia Rating (CDR) scale. The sample of 1,119 participants had a reported dementia prevalence of

6.4%. The sensitivity of CDT was reported as 0.76, and the specificity was 0.81. Of the two reviews[99,120] which only reported pooled study-level data, one review reported pooled data from six studies conducted in "primary care relevant popluations"[99], with a total of 2170 participants. The reported reference standard was DSM criteria and the National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria. Accuracy scores were reported as ranges, with sensitivity between 0.67-0.98 (95% CI 0.39-1.00) and specificity from 0.69-0.94 (95% CI 0.54-0.97). The other aggregated review[120] reported study data pooled across two different scoring systems: CDT Shulman[270] and CDT Sunderland[271]. For CDT Shulman with a pooled study population of 2096 (1266 with dementia), reported sensitivity was 0.893 (95% CI 0.75-0.89) specificity 0.84 (95% CI 0.69-0.92). CDT Sunderland had a pooled study population of 1757 (528 people with dementia) and reported sensitivity of 0.76 (95% CI 0.69-0.83) specificity 0.85 (95% CI 0.76-0.91). Whilst individual study details were not available within the review, the high number of people with dementia within the study samples suggest they may not have been representative of a general practice population.

Taking the systematic review evidence in total, the optimum threshold for CDT in detecting cognitive impairment was unclear from the data presented.

*Mini-Cog accuracy*

The Mini-Cog also featured as an index test in four of the included systematic reviews[120,125,141,236], with one study[224] featured in three reviews[125,141,236] and the other five studies[144,254-257] featuring in one review each.

Of the three reviews reporting the findings of one study[224], one review[236] stated the setting was primary care with a dementia prevalence of 6% in a study population of 1.179. In this review, Mini-Cog sensitivity was reported as 0.76 (95% CI 0.65-0.85) and specificity as 0.89 (95% CI 0.87-0.91) compared against a reference standard of a clinical diagnosis using DSM-III-R and NINCDS-ADRDA criteria. These findings corresponded with those of the second review[125] which reported the study population of 1,179 with 6.4% dementia prevalence – although this review also stated that the same study was based upon a random community sample rather than primary care. This review reported sensitivity as 0.76 and specificity as 0.89 compared against a reference standard using CERAD, DSM-IV and NINCDS-ADRDA criteria. The third review[141] stated that the same study took place in a community setting, with a study

population of 1,119. This review did not report any accuracy data but did state that compared to MMSE at a threshold of 24, Mini-Cog was less sensitive than MMSE whereas at an MMSE cut off of 25, Mini-Cog was more sensitive but less specific than MMSE. The stated reference standard in this review was MMSE and a standardised neuropsychological battery incorporating DSM IIIR and NINCDS-ADRDA criteria.

Two other studies[144,254] reported in one review[125] reportedly drew on community samples yet reported dementia prevalence of 52% and 62% respectively, suggesting these were highly selective and atypical populations. One study[144] reported sensitivity of 0.99 and specificity of 0.93 compared to a reference standard of CAMDEX and DSM-IV in a sample of 249 people with a mean age of 74 years and with 50% non-native English speakers. The other study[254] only featured in this review reported sensitivity of 0.84 and specificity of 0.81 compared to a reference standard of the Cognitive Abilities Screening Instrument (CASI) in a sample of 371 people (mean age =75 years old, 64% non-English speaking).

One other review[120] reported pooled study data across different brief cognitive assessments, with nine studies reporting accuracy data on Mini-Cog and three of those studies drawn from a primary care population. Three studies set in primary care[255-257] (identified by their referenced titles) reported accuracy data although individual reference standards were not reported. Instead, pooled reference standards were reported including DSM, ICD, NINCDS-ADRDA, CAMDEX, GMS-AGECAT, CERAD and CDR. One study reported sensitivity of 0.78 (95% CI 0.71-0.84) and specificity of 0.59 (95% CI 0.50-0.67), one reported sensitivity of 1.00 (95% CI 0.84-1.00) and specificity of 0.85 (95% CI 0.81-0.89), and one reported sensitivity of 0.80 (95% CI 0.56-0.94) and specificity of 0.73 (95% CI 0.69-0.77).

It is difficult to draw meaningful conclusions from the results presented within this last review without knowing individual study data such as population size and setting, individual reference standards used and other demographic details such as age, gender and education levels. These data are not presented within the review and cannot be extracted from the primary study as only pooled primary data were reported within the review.

*IQCODE accuracy*

The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) is a 26-item questionnaire administered to an informant which asks about changes in daily cognitive function of the person being assessed. There is also a shorter version (the IQCODE-short form) of 16 questions. Of the three reviews[99,120,261] which featured the IQCODE, only one [120] reported studies using the short and long versions, and this review only reported pooled study level data across studies in populations including clinic, hospital, community, primary care and 'others'. The other two reviews[99,261] both reported data for the long (26 questions) IQCODE.

One review [261] reported data from a single cross-sectional study using the long IQCODE compared to clinical diagnosis in order to identify dementia within a primary care setting. The study assessed 230 people recruited from general practice in Hawaii with a dementia prevalence of 7%. The majority of respondents were older women with limited higher education. Within this study, IQCODE accuracy was assessed at various thresholds, with a balance made between sensitivity and specificity. At an IQCODE threshold of 3.2 sensitivity was 1.00, and specificity was 0.76. At the 3.7 threshold, IQCODE sensitivity was 0.75 and specificity was 0.98.

One review[99] reported pooled study data from 5 studies with a total of 1108 participants. The review authors reported that these studies were conducted in "primary care relevant populations" but no further detail was given on this, nor were individual study level data reported to allow further scrutiny. The reference standard used was DSM criteria and the NINCDS-ADRDA. As the latter measure is specifically focused towards identifying Alzheimer's disease, this suggests that at least one of the aggregated populations did not reflect primary care. The review reported IQCODE ranges of sensitivity = 0.75–0.88 (95% CI 0.41–1.00) and specificity = 0.65–0.91 (95% CI 0.59–1.00),. With the cut points not reported and an administration time of $\leqslant$20 minutes given. Again, this indicates a lack of relevance for a primary care population.

The third review[120] also reported pooled study data for 7 study cohorts using the short IQCODE and for 17 studies using the long IQCODE. Within these studies, only one in each represented a primary care population, and as only aggregated data were reported these results are of limited relevance and applicability to our overview. For the short IQCODE compared against a large number of reference standards consisting of DSM, ICD, NINCDS-ADRDA, CAMDEX, GMS-AGECAT, CERAD, CDR, sensitivity

was 0.89 (95% CI 0.85-0.92) and specificity was 0.82 (95% CI 0.63-0.93). The thresholds used were also not reported, and again this contributes to the limited usefulness of these results.

As with the Mini-Cog accuracy reported above, It is not possible to draw meaningful conclusions from the results presented within these two reviews reporting pooled study data without knowing individual study details such as test thresholds, population size and setting, individual reference standards used and other demographic details such as age, gender and education levels. These data are not presented within the review and, as with the Mini-Cog, cannot be extracted from the primary study as only pooled primary data were reported within the review.

### 4.5.3. Direct comparisons between brief cognitive assessments within the overview

Within this overview, only one review[125] reported findings from a study[142] directly comparing the diagnostic performance of GPCOG, MMSE (at a threshold of <25) and AMT (at a threshold of <8) within the same community sample (N=283). Participants were aged over 50, with a mean age of 80 years old. Dementia prevalence was 29%, and the setting was reported as general practice and a combined reference standard was used consisting of CAMDEX and DSM-IV criteria. The type of GPCOG measure was not reported in the review, but on checking the original study the authors stated that they used GPCOG two stage.

### 4.5.4. The importance of direct comparisons

As study design is such an important factor in determining the sensitivity and specificity of brief assessments, comparing these measures across different studies is of limited diagnostic value and may well provide misleading interpretations. The limited diagnostic accuracy evidence from direct comparisons of brief cognitive assessments was a clear deficit in seeking to understand which tool may be best suited for general practice as part of a clinical evaluation for possible dementia.

The next logical phase for investigation was to carry out a systematic review of evidence for direct comparisons of brief cognitive assessments, using the two most frequently-assessed and reasonably comparable tools identified within this overview – the MMSE and GPCOG. This systematic review of direct comparisons would allow the possibility to account for variation not due to differences in study design, and to explore systematically which of these tools performed better within a general practice

population. The systematic review and discussion of findings are reported in detail in the following chapter.

### 4.5.5. Analytical and clinical performance measures

As expected, there was no analytical performance evidence presented within the thirteen systematic reviews included within the overview. As this thesis focuses explicitly on the diagnostic accuracy of brief cognitive assessments, It would be expected that clinical performance measures should feature strongly throughout the evidence from the overview. Diagnostic accuracy was measured and reported across all included systematic reviews, as this was part of the inclusion criteria in this overview. Other measures were less clearly present across the included evidence, and the patterns of how they feature within the overview are detailed in Table 12.

Table 12 Overview-identified evidence from systematic reviews of factors beyond diagnostic test accuracy in indirect comparisons of GPCOG and MMSE

| | **Systematic reviews** | **Brodaty 2006** | | **Mitchell 2010a** | | **Tsoi 2015** | |
|---|---|---|---|---|---|---|---|
| | **Index tests** | GPCOG | MMSE | GPCOG | MMSE | GPCOG | MMSE |
| **Clinical performance** | Inter-rater reliability | X | ? | | | | |
| | Test-Retest reliability | X | X | | | | |
| **Clinical effectiveness** | Face validity | X | X | | | | |
| | Internal consistency | X | x | | | | |
| | Education Bias | X | X | | | | |
| | Language/ Culture Bias | ? | X | | | | |
| **Broader impact** | Time taken (mins) | 4.5 | 5-10 | 4.5 | 9-15 | ≤10 | 6-10 |
| | Ease of admin | | | X | X | | |
| | Practicability for general practice | | | X | X | | |
| | Clinician preference | | | X | X | | |
| | Acceptable to patients | | | X | X | | |

x , Reported/measured; blank, not reported/ measured; ? unclear

All 13 included systematic reviews reported diagnostic test accuracy measures of sensitivity and specificity and combined accuracy scores. Two systematic reviews[222,236] also reported positive and negative predictive value and misclassification rates. As shown in Table 12, one review[236] reported interrater and test-retest reliability data on GPCOG and MMSE, although there were insufficient data on the interrater reliability of MMSE as reported in the systematic review (i.e. it was noted, but no numerical data were presented).

All three systematic reviews that indirectly compared the performance of MMSE and GPCOG[120,141,236] reported assessments of test administration time (i.e. the length of time of takes to carry out a test) as shown in Table 14. None of these reviews reported how these timings were reached, such as whether they were measured from the point the test administrator began to explain the testing process to the point the test was complete, results explained and patient de-briefed. Administration time for GPCOG was consistent across two reviews[141,236] at 4.5 minutes, and in one review[120] was reported as taking less than 10 minutes. MMSE administration time was reported as a range in all three systematic reviews and was less consistent, with one review[236]reporting an administration time range between 5-10 minutes, one review[120] reporting a range of 6-10 minutes and one review[141] reporting a range of 9-15 minutes. All included reviews made some statement within their discussion section based upon administration time.

One systematic review[141] reported assessments on ease of administration, practicability for general practice, clinician preference and acceptability to patients (potentially assessed by GP report, although unclear) for both the MMSE and GPCOG. Neither of the other systematic reviews reported on these or other broader impact measures.

### Implications for research

The purpose of an overview must be crystal clear. The paradox is that overviews intuitively seem like a good idea. The practical experience developed here is that one needs the findings from all included reviews to line up in order for the overview to usefully function; however, if all the evidence agrees in one direction, then why is an overview needed? Bearing this fundamental point in mind, anyone contemplating an overview must be able to clearly answer these two questions: in what circumstances would an overview a) work and b) be useful?

Quality issues in the conduct of primary studies and systematic reviews have already been addressed within this chapter and the wider thesis, but this was another stark finding within the overview. Poorly-conducted studies were ill-suited to answer the questions they claimed to address, and similarly poorly-conducted systematic reviews amplified this problem. Related to this issue, it was found that weak systematic review data did not fairly reflect the primary research that does exist.

Recent growth in the popularity of overviews[272], has not yet impelled clear guidance on conducting overviews of diagnostic test accuracy - although this is slowly being addressed by researchers within the field of overview methods development[189,273-275].

Finally, there are currently no quality assessment tools for overviews of diagnostic test accuracy. The recently-developed PRISMA-DTA reporting guidelines[77] will only help to improve assessment of reporting quality, and consequently may also improve conduct and quality assessment in evidence syntheses of diagnostic test accuracy data.

## 4.6. Summary of suggested improvements

Whilst many methodological challenges have been addressed within this chapter, two prominent challenges remain in the conduct of overviews – broadly, and specifically in relation to overviews of diagnostic accuracy – and in methods used for direct comparisons of test data.

Overviews are a relatively new methodological approach and consequently a number of aspects of overview methodology remain uncertain. It is the responsibility of a research team to decide on their approach before conducting an overview; central to this is determining what type of overview is to be conducted. Clear decisions relating to the research questions and objectives to be addressed by the overview are a fundamental first step during the initial planning stages for an overview, and should be developed with the involvement of key stakeholders. Following best practice, these aspects should be covered within a published overview protocol as a mechanism for ensuring transparency and reducing opportunities for introduction of bias in the conduct of the overview[189].

Despite a need for improved guidance for the conduct of overviews[189], there are a number of resources available which support the conduct of overviews[188,189,191,276], and updates to the relevant chapter of the Cochrane Handbook are currently in production[191]. Further guidance on the less common types of overview (such as those addressing reviews of diagnostic tests accuracy and prognosis) and more challenging aspects of overview production, such as methods for narratively synthesising findings, dealing with missing data, poor reporting, and dealing with complexity versus granularity [277] would be a great benefit to those tackling overviews.

Small numbers of carefully designed comparative studies offer the opportunity to resolve questions about whether one test is better than another. Such questions can remain unclear even after many primary accuracy studies focussing on individual tests have been conducted and reviewed.

Authors of comparative accuracy studies need to carefully choose and justify the comparisons they make. Comparisons of convenience which have little clinical importance may simply compound difficulties making sense of traditional single arm accuracy studies which are often profuse yet may have limited clinical application.

Comparative accuracy studies need careful interpretation. Sensitivities and specificities should not be interpreted in isolation; ROC curves provide additional important information. They may also increase the number of studies which can contribute to a review as the ROC curve is derived by measuring accuracy across a range of thresholds.

Systematic reviews of ROC curves are still fairly uncommon. This may indicate a need for methodological development to make it easier to review and meta-analyse accuracy studies which produce ROC curves, or it may be a lack of available ROC data – although this is unlikely, given the experience in this review where ROC curves were produced in all five included studies.

Quality appraisal of comparative accuracy studies remains in its infancy, yet there are some specific aspects of quality and reporting (such as assessing more than one index test and unequal reporting of key index information) as well as issues of applicability and bias discussed above, which would make tailored tools very useful.

There is an opportunity for comparative accuracy studies to increase their explanatory value by providing information about basic functions of testing such as discordance (i.e. why do certain people test positive on one test and not the other when disease is present?), test administration time, acceptability of the test to the patient and the clinician, and the interaction between disease severity and test performance. Whilst some of these aspects would require additional targeted research, other factors such as administration time may require little additional data collection and simply call on further consideration during the analytical phase.

The current guidance on the reporting of diagnostic test accuracy reviews from the PRISMA-DTA Working Group[77] points to progress in diagnostic test accuracy

methodology development. This progress needs reinforcement by educators and regulators to continue this trajectory, and further guidelines are needed on the conduct and reporting of diagnostic accuracy information at the study level as well as in the conduct of systematic reviews and overviews.

By anchoring this discussion of methodological challenges in real examples encountered during this PhD research, it is hoped that this has provided clear illustrations of the complex issues inherent in this area and offered practical solutions to these problems.

### 4.6.1. Implications for clinical practice

The lack of clear evidence on the most suitable brief cognitive assessments for use in general practice is a clear barrier to dementia diagnosis in primary care not addressed within recent clinical guidelines [173]or by current campaigns to improve diagnosis rates.

To return to the points made at the start of the discussion, there are three main difficulties for those in clinical practice looking for evidence-based guidance on the most accurate cognitive assessment tools for use in general practice. These are: a lack of suitable primary data on which cognitive tests are most appropriate for use in general practice; poor quality primary evidence which adds little to current understanding; and poor quality secondary analysis of existing data which adds to the general confusion between clinical guidelines and official policy.

This overview has demonstrated that within current evidence there is not one brief cognitive assessment that clearly emerges as superior to others in terms of test accuracy. The breadth of diagnostic test accuracy evidence is mixed, and there is a great deal of variation in reported levels of sensitivity and specificity. The selected threshold or cut-point significantly influences the dynamic between these interrelated scores but again, within this overview was found to be under-reported and under-recognised within review data – and in some cases the threshold was not prespecified [278] allowing the 'best threshold' to be chosen post hoc. Population prevalence also has a substantial influence and in the majority of studies included within these reviews the prevalence was either not reported or far higher than would be expected within a primary care population; therefore comparisons for general practice should only be drawn with extreme caution and close reference to prevalence data. A useful

conclusion that can be drawn from this overview is that some tools are not suitable for use in general practice.

The MMSE has been popular for many years and is familiar in a clinical setting. Evidence identified within this overview for its diagnostic accuracy in assessing cognitive impairment in general practice is highly mixed; but it is not suitable for use in this population simply due to its lengthy administration time of 7.3±5 minutes[144], longer than practicable for the standard mean general practice appointment length of 10.7 minutes (6.7 standard deviation)[132].

Equally, the IQCODE is reported to have an estimated administration time of 10-15 minutes[248] which renders it unsuitable for clinical use in general practice.

Some of the tools identified within this overview are not sufficiently assessed within a UK general practice-relevant population to be recommended for use in practice. These include the 3 word recall, AMTS, Eurotest, Fototest, BIMCT, John Brown Test, Verbal Fluency, Orientation and PCL.

## 4.7. Overview review team acknowledgements

This overview research was led by the PhD candidate, Harriet Hunt. Dr Elzbieta Kuzma was on the research team in the role of second reviewer. Dr Kuzma contributed to discussions on methodological direction, and assisted with data extraction, quality assessment and analysis. Professor Christopher J Hyde contributed to methodological discussions and provided oversight of the project. Dr Obioha Ukoumunne gave statistical advice to the lead reviewer (HH) which was used to inform the analysis.

Table 13 Summary of evidence and main findings at the level of systematic review

| Review ref. | Reference standard | Index test(s) | T'holds | Setting | *Main findings* | *Other outcomes considered beyond test accuracy* |
|---|---|---|---|---|---|---|
| 204 | NINCDS-ADRDA or DSM or ICD criteria for Alzheimer's disease dementia; McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; and NINDS-AIREN criteria for vascular dementia | MMSE | ≤ 21,<br>≤ 26,<br>≤ 28,<br>≤ 29 | Participants were recruited from: i) secondary care - outpatient clinic (n = 3); ii) secondary care - memory clinics (n = 6) and iii) populational sources (n = 2) | No evidence supporting a substantial role of MMSE as a stand-alone single-administration test in the identification of MCI patients who could develop dementia. Clinicians could prefer to request additional and extensive tests to be sure about the management of these patients. An important aspect to assess in future updates is if conversion to dementia from MCI stages could be predicted better by MMSE changes over time instead of single measurements. It is also important to assess if a set of tests, rather than an isolated one, may be more successful in predicting conversion from MCI to dementia. | - |
| 236 | Clinical diagnosis/clinical diagnosis combined with e.g. DSM-IV | 6-CIT<br>7MS<br>Bowles-Langley Technology Ashford Memory Test<br>MAT<br>RUDAS<br>STMS<br>Time and Change Test<br>SMT<br>CAMCog<br>CDT<br>GPCOG<br>Mini-Cog<br>MIS<br>MMSE<br>SSSI | Various | Validated in "two distinct samples" or Inpatient or Outpatient settings<br>All others validated in general practice, community or population settings | The authors recommend that GPs consider using the GPCOG, Mini-Cog or MIS when screening for cognitive impairment or for case detection. Quick and easy to administer whilst having psychometric properties similar to the MMSE. | Time taken<br>Face validity<br>Internal consistency<br>Education Bias<br>Language/Culture Bias<br>Interrater Reliability<br>Test-Retest Reliability<br>Ease of administration<br>Practicability for general practice<br>*See separate non-DTA data table* |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Short Informant Questionnaire on Cognitive Decline in the Elderly | | | | |
| 220 | diagnosis of dementia according to criteria DSM - IV | Eurotest | - | Mix of secondary (specialist) care and Fase III primary care | Review authors conclude that the Eurotest has adequate diagnostic validity to be used as an instrument to screen for and rule out dementia; negative results are more useful, allowing the presence of dementia to be ruled out with greater certainty. | - |
| 221 | Clinical diagnosis according to DSM-IV-TR | Phototest | - | Primary care | For dementia, Sn was 0.85 (95% CI, 0.82-0.88) and Sp 0.87 (95% CI, 0.85-0.99); Authors' conclusions: the Phototest offers adequate diagnostic accuracy for cogni-tive impairment, and particularly for dementia, that is similar or superior to other instruments widely used in our milieu. Additionally, it is simple, brief, uninfluenced by educational variables and can even be used in individuals who are illiterate. These advantages make it attractive for use in populations with low educational level and/or in time-limited settings such as primary care. | - |
| 205 | Clinical diagnosis defined by DSM, ICD or CRS. | MMSE | 10 cut points 17-26 incl. | Primary care | The authors could not estimate summary DTA in primary care due to insufficient data. "The MMSE contributes to a diagnosis of dementia in low prevalence settings, but should not be used in isolation to confirm or exclude disease. We recommend that future work evaluates the diagnostic accuracy of tests in the context of the diagnostic pathway experienced by the patient and that investigators report how undergoing the MMSE changes patient-relevant outcomes", e.g. time to diagnosis, initiation of treatment/care package, additional testing and place of care. | - |
| 261 | Clinical diagnosis defined by DSM or ICD | IQCODE | 3.6; 3.5; 3.4; 3.3 | General practice/primary care | IQCODE accuracy assessed at various test thresholds, with a "trade-off" between sensitivity and specificity across these cutpoints. At an IQCODE threshold of 3.2 sensitivity: 100%, specificity: 76%; for IQCODE 3.7 sensitivity: 75%, specificity: 98%.<br>Author's conclusions:<br>"It is not possible to give definitive guidance on the test accuracy of IQCODE for the diagnosis of dementia in a primary care setting | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | based on the single study identified. We are surprised by the lack of research using the IQCODE in primary care as this is, arguably, the most appropriate setting for targeted case finding of those with undiagnosed dementia in order to maximise opportunities to intervene and provide support for the individual and their carers". | |
| 222 | Clinical diagnosis using DSM III, IIIR or IV | MMSE | 21v22, 22v23, 23v24, 24v25, 25v26, 26v27 28v29 | Primary care | In those studies conducted purely in primary care the Se, Sp, PPV and NPV were 78.4%, 87.8%. 53.6% and 95.7%, respectively. MMSE offers modest accuracy with best value for ruling-out a diagnosis of dementia in community and primary care. For all other used it should be combined with or replaced by other methods. | - |
| 141 | Clinical diagnosis using DSM/IPA-WHO criteria/GMS-AGECAT 4 or 5+ CAMDEX | MMSE, GPCOG, AMTS, Mini-Cog PCL, SPMSQ, MSQ, BIMT | NR | Primary care | The authors conclude that for a primary care setting (where prevalence is usually modest) the best individual tools were the AMTS/MSQ (combined, the MSQ and the PCL. The AMTS was preferable to the MMSE for case-finding, whereas the SMPSQ was inferior. For screening, the MMSE was optimal and…the best tool for primary care physicians who want a rule in and rule out tool, **if** length is not a major consideration.<br>"at least 30 well-studied alternatives to the MMSE exist and several seem to be briefer but no less accurate than the MMSE…In primary care…these methods would help detect on 13 of 20 possible cases but rule out 18 of 20 healthy individuals (overall correct in 17 out of 20) suggesting these brief batteries may be ideally suited to a first step screen followed by a more comprehensive in those who screen positive" (p. 779). | Comparative accuracy conducted with 5 instruments (AMTS/MSQ, MSQ, WINDSET, PCL and AMTS).<br>Time taken given in all cases.<br>*See separate non-DTA data tables* |
| 231 | MiniCog and DSM criteria/DSM criteria and GMS-AGECAT/DSM criteria and NINCDS-ADRDA/DSM | CDT, 3 word recall, John Brown test, Verbal fluency (animals), orientation | NR | Primary care | Across 9 analyses involving 4,875 individuals, the prevalence of dementia was 10.6%. The pooled sensitivity was 68.5%, "corrected to 69.5% (95% CI = 62.1% - 76.4%)" and the pooled specificity was 85.9%, "corrected to 82.5% (95% CI 74.1%-89.5%) PPV 36.5% and NPV 95.8%" (p.792).<br>8 methods were compared directly to the MMSE all performed worse. "A fraction correct pooled relative risk was reported of 0.911 (95% CI = 0.865-0.959) $x^2$ = 12.7 (df=1) p.0.0004. Reported for both Sensitivity (pooled relative risk = 1.107, 95% CI = 0.902-1.360) $x^2$ = 0.958 (df=1) p=0.33 and Specificity (pooled relative risk = 0.872, 95% CI=0.807-0.942, $x^2$ = 12.072421 $(df = 1)p = 0.0005)$." The Clock Drawing Test was "subject to multiple | Comparative accuracy was conducted with 5 instruments.<br>Time taken was given in all instances.<br>*See separate non-DTA data tables* |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | independent testing in primary care" but demonstrated no superiority to MMSE (p.793).<br>***HH N.B. I could not see reported MMSE scores within the review publication, so the above conclusions are reported direct from review authors.*** | |
| 148 | DSM-IV criteria | RUDAS | < 23 | For clinicians to use in culturally and linguistically diverse populations | "The RUDAS was assessed in 1236 participants and was found to have a pooled sensitivity of 77.2% (95% confidence interval [CI] 67.4–84.5) and a pooled specificity of 85.9% (95% CI 74.8–92.6) yielding a positive likelihood ratio of 5.5 (95% CI 2.9–10.7) and a negative likelihood ratio of 0.27 (95% CI 0.17– 0.40)<br>A pooled estimate of the correlation between the RUDAS and the Mini-Mental State Examination (MMSE) was 0.77 (95% CI 0.72– 0.81) and significant heterogeneity ($I^2$ = 63.1%). Results of the RUDAS were less affected by language and education level than the MMSE." (p.E169) | Comparative accuracy assessed against MMSE as reported in box to left.<br>Education level, effect of language, immigrant status, clinician preference, test-retest reliability and interrater reliability assessed in some studies. *See separate non-DTA data table* |
| 125 | CAMDEX, DSM-IV | GPCOG<br>AMT<br>MMSE | <8<br><25 | General practice | Review authors concluded that the GPCOG outperformed the MMSE in a general practice sample. "However, the composition of the test was changed post hoc, and the available data are incomplete" [??]<br>"…the limited cognitive domains that all these tests assess may make them prone to miss non-Alzheimer's dementias" (p.475) | Other assessment domains (language, attention, memory, visuospatial skills, executive function,) and monitoring disease progression discussed. |
| 99 | DSM criteria and NINCDS-ADRDA | MMSE,<br>3 word memory test,<br>6-item screener,<br>7MS,<br>AMT,<br>Benton's Orientation Test,<br>CDT,<br>Cognitive Assessment | Various | Primary care and community (primary care-relevant populations) | Review authors reported that "only 12 brief instruments have been studied more than once in well-designed diagnostic accuracy studies that evaluated their ability to detect dementia in primary care–relevant populations: the MMSE (*k* = 25; *n* = 12 348), the Clock Drawing Test (CDT) (*k* = 7; *n* = 2509), verbal or category fluency tests (*k* = 6; *n* = 2083), the short or full Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (*k* = 5; *n* = 1108), the Memory Impairment Screen (MIS) (MIS: *k* = 4; *n* = 1671; MIS by telephone: *k* = 1; *n* = 300), Mini-Cog (*k* = 4; *n* = 1570), the Abbreviated Mental Test (AMT) (*k* = 4; *n* = 824), the Short Portable Mental Status Questionnaire | Other assessments of benefits and harms, caregiver interventions, non-pharma I'ventions aimed at the patients. *See separate non-DTA data table*. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Screening Test, Free and Cued Selective Reminding Test, Functional Activities Q'airre, GPCOG, Hopkins Verbal Learning Test, Immediate Recall, IQCODE (short), IQCODE full, Katz ADL, Kendrick Cognitive Tests, Labyrinth Test, Memory Function 2, MIS, Memory Impairment Screen-Telephone, MSQ, MiniCog, MMSE, Minimum Data Set Cognition Scale, MMblind, Oral Traits, Orientation | | | (SPMSQ) (*k* = 4; *n* = 1057), the Mental Status Questionnaire (*k* = 2; *n* = 522), the Free and Cued Selective Reminding Test (FCSRT) (*k* = 2; *n* = 734), the 7-Minute Screen (7MS) (*k* =2; *n* =553), and the Telephone Interview for Cognitive Status (TICS) (*k* =2; *n* =677)" (p. 604).<br>…"The best-studied instrument was the MMSE. Pooled estimates across 14 studies (*n* _ 10 185) resulted in sensitivity of 88.3% (95% CI, 81.3% to 92.9%) and specificity of 86.2% (CI, 81.8% to 89.7%) for the most commonly reported cut points of 23/24 or 24/25. The CDT, Mini-Cog, MIS, SPMSQ, AMT, FCSRT, 7MS, TICS, and IQCODE can also have acceptable test performance; however, less evidence supported the use of each of these instruments and had limited reproducibility in primary care– relevant populations and unknown optimum cut points for each instrument. The CDT had a wider range of sensitivity and specificity (67% to 97.9% and 69% to 94.2%, respectively), and the optimum cut point is unclear from the body of literature we examined. The Mini-Cog probably has better sensitivity than the CDT alone (76% to 100%) but with a possible tradeoff of lower specificity (54% to 85.2%). Although the MIS can have relatively good test performance in screening for dementia (sensitivity, 43% to 86%; specificity, 93% to 97%), the sensitivities in the 2 good-quality studies (*n* _ 948) were low (about 40%). Likewise, the AMT can have relatively good test performance in screening for dementia (sensitivity, 42% to 100%; specificity, 83% to 95.4%), but 1 fair-quality study (*n* _ 289) had low sensitivity (42%) and no studies were done in the United States. The SPMSQ, FCSRT, 7MS, and TICS also have reasonable test performance, but this is based on a limited number of studies. The verbal fluency tests had worse performance than other instruments regardless of cut point. The IQCODE, a self-administered informant-based screening tool, had a sensitivity of 75% to 87.6% and a specificity of 65% to 91.1%. The 6-Item Screener, Visual Association Test, General Practitioner Assessment of Cognition, ADL/IADL, Benton Orientation Test, Delayed Recall Test, and Short Concord Informant Dementia Scale all had greater than 80% sensitivity and specificity to detect dementia in a single study, but their test performance has | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Concentration Memory, Rey figure copy, Self-Administered Gero-cognitive Examination, SBT, Short Concord Informant Dementia Scale, SPMSQ, Storandt Battery, Subjective Menory Impairment, Sweet 16, TICS, Trailmaking A and B, Verbal fluency, Visual Association (VAT), Word List Learning | | | not been reproduced in other primary care–relevant populations" (p. 605).<br>**Review authors' conclusions**:<br>"…only a handful of instruments have been studied in more than 1 study applicable to primary care. Although the MMSE is the best-studied instrument, it has the longest administration time and is not available for public use without cost. Other publicly available instruments that have been studied in primary care–relevant populations can have adequate test performance, including the CDT, Mini-Cog, MIS, AMT, SPMSQ, FCSRT, 7MS, and IQCODE. However, the AMT, SPMSQ, FCSRT, and 7MS have limited evidence, and each has been studied only once in English. Although other instruments seem to have adequate test performance (such as the 6-Item Screener, Visual Association Test, General Practitioner Assessment of Cognition, ADL/IADL, Benton Orientation Test, Delayed Recall Test, and Short Concord Informant Dementia Scale), each of them has been studied only once in primary care–relevant populations." | |
| [120] | DSM, ICD, NINCDS-ADRDA, CAMDEX, GMS-AGECAT, CERAD, CDR (not reported individually) | CDT, MiniCog, MIS, Verbal fluency test, AMT, GPCOG, MMSE, MoCA (query – not under 10 | NR | Primary care | From the authors – **accuracy of MMSE:**<br>With different cut-off threshold values, we found considerable variation in the sensitivity and specificity estimates reported by individual studies. The sensitivities ranged from0.25 to 1.00, and the specificities ranged from 0.54 to 1.00. The heterogeneity among studies was large, with *I*2 statistics for sensitivity and specificity of 92% and 94%, respectively. The diagnostic accuracy is summarized by meta-analysis (Table 3). The combined data in the bivariate random-effects model gave a summary point with 0.81 sensitivity (95% CI, 0.78-0.84) and 0.89 specificity | Administration time, number of questions, components of screening tests<br>*See separate non-DTA data table* |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | mins admin time) | | | (95% CI, 0.87-0.91). The HSROC curve was plotted with a diagnostic odds ratio of 35.4, and the AUC was 92%(95%CI,90%-94%).<br>**Pooled accuracy of tests:**<br>MMSE reported pooled sensitivity of 0.81 (0.78-0.84) and pooled specificity of 0.89 (0.87-0.91), positive pooled LR of 7.45 (6.25-8.88) and negative pooled LR of 0.21 (0.18-0.25).<br>CDT (Shulman) reported pooled sensitivity of 0.83 (0.75-0.89) and pooled specificity of 0.84 (0.69-0.92), pooled positive LR 5.02 (2.61-9.64) and pooled negative LR 0.20 (0.14-0.29). CDT (Sunderland) reported pooled sensitivity 0.76 (0.69-0.83) and pooled specificity of 0.85 (0.76-0.91), pooled positive LR 5.09 (3.18-8.13) and pooled negative LR 0.28 (0.20-0.38). MiniCog reported pooled sensitivity 0.91 (0.80-0.96) and pooled specificity 0.86 (0.74-0.93), pooled +ve LR 6.56 (3.25-13.24) and poole –ve LR 0.10 (0.04-0.25).<br>MIS reported sensitivity 0.797 (0.68-0.86) and specificity 0.91 (0.84-0.96), VF test sensitivity 0.80 (0.73-0.86) and specificity 0.82 (0.73-0.88), AMT sensitivity 0.88 (0.82-0.92) and specificity 0.85 (0.81-0.89), GPCOG sensitivity 0.92 (0.81-0.97) and specificity 0.87 (0.83-0.90; MoCA sensitivity 0.91 (0.84-0.95) and specificity 0.81 (0.71-0.88).<br>“All tests presented with AUCs of at least 85%, and most of the tests had comparable performance to that of the MMSE. The Mini-Cog test and the ACE-R were the best alternative tests. Among the studies with the Mini-Cog test,10,34-41 the pooled sensitivity was 0.91 (95% CI, 0.80-0.96), and the pooled specificity was 0.86 (95% CI,0.74-0.93)” (p. E6). | |

## Appendices

Appendix 1 Overview protocol published on PROSPERO 02/06/2015

Published on PROSPERO 02 June 2015 ref. CRD42015022078

# A review of existing systematic reviews summarising the accuracy of brief cognitive assessments for identifying dementia, particularly for use in primary care [protocol]

### Introduction

Improved dementia diagnosis is a priority of the UK Government's National Dementia Challenge, as well as a key focus of the World Health Organisation and the G8. The Alzheimer's Society's recent report *Dementia 2014: Opportunity for change* (Alzheimer's Society, 2014) highlights the urgent need to create a diagnostic pathway which takes people fluently from first presentation to their GP through to memory clinics and long-term support mechanisms following a diagnosis. It is therefore crucial to address delays in the early stages in order to increase access to care and support for the individual and the people around them.

GPs are normally the first point of contact for patients and carers concerned about possible dementia, though general practice probably under-diagnoses the condition (Iliffe et al., 2009; Connolly et al., 2011). There are a number of possible reasons for this under-diagnosis. Many GPs report a lack of certainty in using assessment tools alongside concerns around the consequences of misdiagnosing dementia (Aminzadeh et al., 2012, Bradford et al., 2009; Cahill et al., 2006; Iliffe et al., 2003; Koch & Iliffe, 2010; Sarkar et al., 2012;). It is unclear which brief cognitive assessments for dementia would be best for use in primary care, and how accurate they are. In the UK, there is a lack of agreement between leading organisations on which tests should be used for dementia identification in primary care. For example, the National Institute for Health and Care Excellence (NICE) recommend using the Mini Mental State Examination (MMSE), 6-Item Cognitive Impairment Test (6-CIT), General practitioner assessment of cognitive function (GPCOG) or the 7-Minute Screen, and the Alzheimer's Society who recommend using the Abbreviated mental test score (AMTS), GPCOG and the Mini-cog (Ballard, 2015; NICE Pathways, 2012;). The accuracy of many of the commonly-used brief cognitive assessments for dementia is imperfect.

In building the evidence base, a number of Cochrane Reviews have explored the individual value of tests for dementia to general practitioners (Harrison et al., 2014), to secondary care units such as memory clinics (Quinn et al., 2013a), and in community screening (Fage et al., 2013; Harrison et al., 2015) - or across a number of these settings (Aravalo-Rodriguez et al., 2013; Creavin et al., 2014; Davis et al., 2013; Hendry et al., 2014).

These Cochrane reviews and other systematic reviews have explored the diagnostic test accuracy of tests in isolation, and across a broad range of populations and settings. The diversity of systematic review evidence now available would benefit from being summarised within an overview review, which is the focus of this protocol.

This overview of reviews is positioned within a broader programme of research addressing the question "how can we optimize cognitive assessment in primary care to support a more effective route to dementia diagnosis?", and aims to summarise the existing systematic review evidence for the diagnostic test accuracy of brief cognitive assessments (in addition to informant ratings where used), focussing particularly on their use in a primary care setting.

### Review question

What is the existing systematic review evidence for the accuracy of brief cognitive assessments, plus informant ratings where used, for identifying dementia, particularly for use in primary care?

**Methods**

An overview of reviews protocol with clear description of inclusion and exclusion criteria for reviews and a detailed search strategy will be published on the International Prospective Register of Systematic Reviews (PROSPERO) in advance of the overview review being conducted. Data collected will be from included systematic reviews, and limitations of the included reviews will be assessed alongside quality of evidence based – where data is available – on assessments reported in the included systematic reviews.

**Searches**

We will search the following databases for eligible systematic reviews:

- Cochrane Database of Systematic Reviews (CDSR)- Medline (Ovid),EMBASE, PsychINFO

- EMANATE (Dementia Meta-Evidence) database, created by University of Exeter Medical School researchers

Reviewers will contact key authors and conduct hand searching of reference lists from included and review articles.

Details of the search strategy are shown in the Appendix.

**Types of study to be included**

We will include peer-reviewed systematic reviews and meta-analyses (including Cochrane Reviews) of studies investigating the diagnostic test accuracy of any brief cognitive assessment for dementia within primary care or general practice as assessed by a healthcare professional. We define a brief cognitive assessment as a discrete tool designed to be used to conduct a quick structured evaluation of the patient directly (e.g. MiniCog, MMSE) or as an informant assessment of the patient (e.g. short IQCODE, GPCOG).

**Condition being studied**

All-cause dementia and key dementia subtypes including Alzheimer's disease, vascular dementia, dementia with Lewy bodies, frontotemporal dementia and mixed dementia.

**Participant/population**

Inclusion criteria:

- systematic reviews and meta-analyses of test accuracy evidence including Cochrane Reviews
- adult (>18 years old) participants assessed for dementia (all-cause dementia and key dementia subtypes including Alzheimer's disease, vascular dementia, dementia with Lewy bodies, fronto-temporal dementia and mixed dementia) at any stage of progression
- using brief cognitive assessment tools, plus informant ratings where used,
- Within primary care as assessed by a healthcare professional.
- Updated reviews will be included, with the most recent version taking precedence.
- Conference abstracts will be included and we will attempt to contact all relevant corresponding authors for full text versions or available test accuracy data. If this detail is not available or authors cannot be contacted, this data will be retained and reported within the review.

Brief cognitive assessments will be limited to those that take up to 10 minutes to conduct. Our analysis will include the Mini Mental State Examination (MMSE); whilst the timings for

this test are contested, it is one of the most commonly used tests and often characterised as taking 'up to' 10 minutes (Cordell et al., 2013; RCP, 2012; Ballard et al., 2015). Brief cognitive assessments plus informant ratings will be included where used. A list of all included tests will be presented within the overview review.

Exclusion criteria:

- We will exclude systematic reviews of assessment tools that are not explicitly brief cognitive assessments (i.e. that take longer than 10 minutes to conduct, and are not explicitly focussed on the assessment of cognitive performance).
- Reviews that do not contain accuracy (sensitivity and specificity) data will be excluded.
- We will also exclude systematic reviews of studies conducted in environments other than primary care (e.g. specialist memory clinics), and any systematic reviews of evidence for the accuracy of combined brief cognitive assessments for dementia in primary care (although evidence will be retained for future inquiry).
- Assessments must be conducted within primary care or general practice and conducted by a healthcare professional.

**Data extraction, selection and coding**

All data will be managed using the latest version of EndNote software. The first 15 sources will be pilot title and abstract screened by two reviewers (HH & EK) according to the inclusion and exclusion criteria. Subsequent discussion will inform the screening notes. Title/abstract screening and full text screening will be conducted by the same two reviewers. A third reviewer (CH) will resolve any disagreements.

A bespoke abstraction form will be piloted by two reviewers (HH & EK) using two sources. Key data extracted will include characteristics of included systematic reviews (references and author details, overall goal of review, date review conducted, date published, participant details, included study details such as authors, year of study, date of publication, country of study, outcomes reported, test timings and general review limitations as well as test accuracy data. The form will be accompanied by a briefing document explaining how it should be used. Data will be abstracted by one reviewer (HH) and spot-checked by a second (EK), with a third reviewer (CH) providing moderation as required.

**Risk of bias (quality) assessment**

There is currently no clear guidance of the quality assessment for overview reviews of diagnostic test accuracy studies, and evidence of the best quality assessment tools is unclear. The general quality of included systematic reviews will be summarised and notes made on variability of findings across reviews and any important flaws in individual reviews. We will use a checklist approach to assess the quality of systematic reviews based on the AMSTAR checklist (Shea et al., 2007), and will critically evaluate the suitability of this tool in the current application. Results will be presented narratively in the text, and in an appropriate graphic representation of quality assessment.

**Strategy for data synthesis and analysis**

Data will be summarised narratively and using 'Summary of findings' tables. Additional analyses may be possible for comparing across included reviews, in which case we will consult with a statistical specialist regarding the validity and suitability of further analyses.

**Discussion**

Within the Discussion section, we will consider whether the reviews included are sufficient to address all of the objectives of the overview, and if not, we will highlight the gaps in evidence.

As part of this scrutiny, we will consider whether all relevant participants and outcomes have been represented and if not, we will highlight missing evidence. Finally we will assess the relevance of the evidence to the review question, which will lead to an overall judgement of the external validity of the Overview. The context of results of the overview within the context of current practice will be discussed here.

**References**

Alzheimer's Society (2014) Dementia 2014: Opportunity for change. Alzheimer's Society, London. www.Alzheimers.org.uk/dementia2014

Aminzadeh, F., Molnar, F. J., Dalziel, W. B., & Ayotte, D. (2012). A Review of Barriers and Enablers to Diagnosis and Management of Persons with Dementia in Primary Care. Canadian Geriatrics Journal, 15(3), 85.

Arevalo-Rodriguez I, Smailagic N, Ciapponi A, Sanchez-Perez E, Giannakou A, Roqué i Figuls M, Pedraza OL, Bonfill Cosp X, Cullum S. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI) (Protocol). Cochrane Database of Systematic Reviews 2013, Issue 10. Art. No.: CD010783. DOI: 10.1002/14651858.CD010783.

Baker, P. R., Costello, J. T., Dobbins, M., & Waters, E. B. (2014). The benefits and challenges of conducting an overview of systematic reviews in public health: a focus on physical activity. Journal of Public Health, 36(3), 517-521.

Ballard C, Burns A, Corbett A, Livingston G, Rasmussen J (2015) Helping you to assess cognition: A practical toolkit for clinicians. Alzheimer's Society website: http://www.alzheimers.org.uk/site/scripts/download_info.php?downloadID=1045

Benbow, S. M., Jolley, D., & Greaves, I. C. (2015). Improving diagnosis of dementia in primary care. Progress in Neurology and Psychiatry, 19(1), 4-4.

Bradford, A., Kunik, M. E., Schulz, P., Williams, S. P., & Singh, H. (2009). Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. *Alzheimer disease and associated disorders*, *23*(4), 306.

Ballard, C., Burns, A., Corbett, A., Livingston, G., Rasmussen, J. (2015) Helping you to assess cognition: A practical toolkit for clinicians. The Alzheimer's Society. [Accessed online 05/05/2015 http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2532 ]

Cahill, S., Clark, M., Walsh, C., O'Connell, H., & Lawlor, B. (2006). Dementia in primary care: the first survey of Irish general practitioners. *International Journal of Geriatric Psychiatry*, *21*(4), 319-324.

Connolly, A., Gaehl, E., Martin, H., Morris, J., & Purandare, N. (2011). Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging & mental health, 15(8), 978-984.

Cordell, C. B., Borson, S., Boustani, M., Chodosh, J., Reuben, D., Verghese, J., ... & Medicare Detection of Cognitive Impairment Workgroup. (2013). Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimer's & Dementia*, *9*(2), 141-150

Creavin ST, Noel-Storr AH, Smailagic N, Giannakou A, Ewins E, Wisniewski S, Cullum S. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's dementia and other dementias in asymptomatic and previously clinically unevaluated people aged over 65 years

in community and primary care populations (Protocol). Cochrane Database of Systematic Reviews 2014, Issue 6. Art. No.: CD011145. DOI: 10.1002/14651858.CD011145.

Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S. The Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementia disorders (Protocol). Cochrane Database of Systematic Reviews 2013, Issue 10. Art. No.: CD010775. DOI: 10.1002/14651858.CD010775.

Fage BA, Seitz DP, Gill SS, Herrmann N, Smailagic N, Chan CCH, Nikolaou V. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting (Protocol). Cochrane Database of Systematic Reviews 2013, Issue 11. Art. No.: CD010860. DOI: 10.1002/14651858.CD010860.

Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting. Cochrane Database of Systematic Reviews 2014, Issue 7. Art. No.: CD010771. DOI: 10.1002/14651858.CD010771.pub2.

Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. Cochrane Database of Systematic Reviews 2015, Issue 3. Art. No.: CD010772. DOI: 10.1002/14651858.CD010772.pub2.

Hendry K, Lees RA, McShane R, Noel-Storr AH, Stott DJ, Quinn TJ. AD-8 for diagnosis of dementia across a variety of healthcare settings (Protocol). Cochrane Database of Systematic Reviews 2014, Issue 5. Art. No.: CD011121. DOI: 10.1002/14651858.CD011121.

Iliffe, S., Manthorpe, J., & Eden, A. (2003). Sooner or later? Issues in the early diagnosis of dementia in general practice: a qualitative study. *Family Practice*, *20*(4), 376-381.

Iliffe, S., Robinson, L., Brayne, C., Goodman, C., Rait, G., Manthorpe, J. and Ashley, P. (2009), Primary care and dementia: 1. diagnosis, screening and disclosure. Int. J. Geriat. Psychiatry, 24: 895–901. doi: 10.1002/gps.2204

Koch, T., & Iliffe, S. (2010). Rapid appraisal of barriers to the diagnosis and management of patients with dementia in primary care: a systematic review. BMC Family Practice, 11(1), 52.

Larner, A. J. (2015). Speed versus accuracy in cognitive assessment when using CSIs. Progress in Neurology and Psychiatry, 19(1), 21-24.

Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. (2014) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. Cochrane Database of Systematic Reviews 2014, Issue 4. Art. No.: CD010079. DOI: 10.1002/14651858.CD010079.pub2.

RCP (2012) Individual patient outcome measures recommended for use in older people's mental health. *Prepared by the Royal College of Psychiatrists' Faculty of the Psychiatry of Old Age.* Occasional Paper (OP)86. http://www.rcpsych.ac.uk/publications/collegereports/op/op86.aspx

NICE Pathways (2012) Dementia diagnosis and assessment. Developed from Dementia (2006) NICE guideline CG42. [Accessed online 05/05/2015: http://pathways.nice.org.uk/pathways/dementia/dementia-diagnosis-and-assessment#content=view-node%3Anodes-diagnosis-and-assessment]

Sarkar, U., Bonacum, D., Strull, W., Spitzmueller, C., Jin, N., López, A., ... & Singh, H. (2012). Challenges of making a diagnosis in the outpatient setting: a multi-site survey of primary care physicians. BMJ Quality & Safety, 21(8), 641-648.
Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., ... & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology, 7(1), 10.
Smith, V., Devane, D., Begley, C. M., & Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC medical research methodology, 11(1), 15.
Wilcock, J., Jain, P., Griffin, M., Thuné-Boyle, I., Lefford, F., Rapp, D., & Iliffe, S. (2015). Diagnosis and management of dementia in family practice. Aging & mental health, (ahead-of-print), 1-8.

Protocol appendix

Search strategy, formatted for MEDLINE (OVID)

Database: **EMBASE**

Data Parameters: 1974 – 2015 June 09

Date Searched: 10/06/1

| Searches | | Results |
|---|---|---|
| 1 | dementia/ or aids dementia complex/ or alzheimer disease/ or aphasia, primary progressive/ or creutzfeldt-jakob syndrome/ or dementia, vascular/ or diffuse neurofibrillary tangles with calcification/ or frontotemporal lobar degeneration/ or huntington disease/ or kluver-bucy syndrome/ or lewy body disease/ | 221832 |
| 2 | dementia.ti,ab. | 98978 |
| 3 | alzheimer*.ti,ab. | 128772 |
| 4 | 1 or 2 or 3 | 255335 |
| 5 | systematic*.ti,ab. | 308331 |
| 6 | meta-analysis.ti,ab. | 83324 |
| 7 | "systematic review".ti,ab. | 72883 |
| 8 | diagnosis/ or "sensitivity and specificity"/ | 1273396 |
| 9 | 5 or 6 or 7 or 8 | 1615147 |
| 10 | brief cognitive tests.ti,ab. | 66 |
| 11 | cognitive screen*.ti,ab. | 1338 |
| 12 | ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. | 220 |
| 13 | cog*.ti,ab. | 358594 |
| 14 | 10 or 11 or 12 or 13 | 358664 |
| 15 | "primary care".ti,ab. | 98787 |
| 16 | "general practic*".ti,ab. | 40574 |

| | | |
|---|---|---|
| 17 | "GP".ti,ab. | 41203 |
| 18 | 15 or 16 or 17 | 166293 |
| 19 | 4 and 9 and 14 and 18 | 257 |

Search strategy, formatted for MEDLINE (OVID)

Database: **PsychINFO**

Data Parameters: 1806 – 2015 June

Date Searched: 10/06/15

| | | |
|---|---|---|
| 1 | dementia/ or aids dementia complex/ or alzheimer disease/ or aphasia, primary progressive/ or creutzfeldt-jakob syndrome/ or dementia, vascular/ or diffuse neurofibrillary tangles with calcification/ or frontotemporal lobar degeneration/ or huntington disease/ or kluver-bucy syndrome/ or lewy body disease/ | 54710 |
| 2 | dementia.ti,ab. | 46067 |
| 3 | alzheimer*.ti,ab. | 43262 |
| 4 | 1 or 2 or 3 | 73484 |
| 5 | systematic*.ti,ab. | 83565 |
| 6 | meta-analysis.ti,ab. | 15868 |
| 7 | "systematic review".ti,ab. | 11626 |
| 8 | diagnosis/ or "sensitivity and specificity"/ | 35862 |
| 9 | 5 or 6 or 7 or 8 | 130463 |
| 10 | brief cognitive tests.ti,ab. | 34 |
| 11 | cognitive screen*.ti,ab. | 690 |
| 12 | ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. | 131 |
| 13 | cog*.ti,ab. | 341564 |
| 14 | 10 or 11 or 12 or 13 | 341608 |

| | | |
|---|---|---|
| 15 | "primary care".ti,ab. | 21434 |
| 16 | "general practic*".ti,ab. | 4560 |
| 17 | "GP".ti,ab. | 3274 |
| 18 | 15 or 16 or 17 | 26754 |
| 19 | 4 and 9 and 14 and 18 | 86 |

## Appendix 2 Search strategy formatted for the Cochrane Library

Search Name: Dementia overview search

Date Run: 17/08/15

Cochrane Database of Systematic Reviews: Issue 8 of 12, August 2015

All results (320)

Cochrane Reviews (150)

Other reviews (72)

Trials (87)

Methods Studies (10)

Technology Assessments (1)

| ID | Search | Hits |
|---|---|---|
| #1 | MeSH descriptor: [Dementia] explode all trees | 3971 |
| #2 | dementia:ti,ab | 5183 |
| #3 | alzheimer*:ti,ab | 5235 |
| #4 | memory complaint:ti,ab | 47 |
| #5 | #1 or #2 or #3 or #4 | 9325 |
| #6 | systematic*:ti,ab | 31555 |
| #7 | review:ti,ab | 44131 |
| #8 | "meta-analysis":ti,ab | 23014 |
| #9 | metaanalysis:ti,ab | 433 |
| #10 | "test accuracy":ti,ab | 233 |
| #11 | #6 or #7 or #8 or #9 or #10 | 64707 |
| #12 | "brief cognitive test*":ti,ab | 6 |
| #13 | "cognitive screen*":ti,ab | 63 |
| #14 | (screening test* near/2 (dement* or alzheimer*)):ti,ab | 21 |
| #15 | screening:ti,ab | 16059 |
| #16 | (diagnosis or ("sensitivity and specificity")):ti,ab | 24875 |
| #17 | cog*:ti,ab | 27752 |
| #18 | #12 or #13 or #14 or #16 or #17 | 51300 |
| #19 | "primary care":ti,ab | 8628 |
| #20 | "general practic*":ti,ab | 3814 |
| #21 | GP:ti,ab | 2034 |
| #22 | community:ti,ab | 17445 |
| #23 | #19 or #20 or #21 or #22 | 28971 |
| #24 | #5 and #11 and #18 | 320 |
| #25 | #5 and #11 and #18 and #23 | 35 |

N.B. systematic review search filters not used as the Cochrane Reviews database only includes systematic reviews.

Appendix 3 Search strategy, formatted for EMBASE (OVID)

*Embase 1974 to 2015 August 21*

| | | |
|---|---|---|
| 1 | dementia/ or aids dementia complex/ or alzheimer disease/ or aphasia, primary progressive/ or creutzfeldt-jakob syndrome/ or dementia, vascular/ or diffuse neurofibrillary tangles with calcification/ or frontotemporal lobar degeneration/ or huntington disease/ or kluver-bucy syndrome/ or lewy body disease/ | 225474 |

| | | |
|---|---|---|
| 2 | exp dementia/ | 252956 |
| 3 | exp dementia assessment/ | 23039 |
| 4 | exp clinical dementia rating/ | 1011 |
| 5 | exp memory disorder/ | 56295 |
| 6 | dementia.ti,ab. | 101033 |
| 7 | alzheimer*.ti,ab. | 131501 |
| 8 | 1 or 2 or 3 or 4 or 6 or 7 | 292776 |
| 9 | systematic*.ti,ab. | 317342 |
| 10 | meta-analysis.ti,ab. | 87144 |
| 11 | "systematic review".ti,ab. | 76490 |
| 12 | exp review/ | 2082052 |
| 13 | (literature adj3 review$).ti,ab. | 238313 |
| 14 | exp meta analysis/ | 97571 |
| 15 | exp "Systematic Review"/ | 93826 |
| 16 | or/12-15 | 2316420 |
| 17 | (medline or medlars or embase or pubmed or cinahl or amed or psychlit or psyclit or psychinfo or psycinfo or scisearch or cochrane).ti,ab. | 133763 |
| 18 | RETRACTED ARTICLE/ | 7620 |
| 19 | 17 or 18 | 141333 |
| 20 | 16 and 19 | 103832 |
| 21 | (systematic$ adj2 (review$ or overview)).ti,ab. | 95715 |
| 22 | (meta?anal$ or meta anal$ or meta-anal$ or metaanal$ or metanal$).ti,ab. | 105476 |
| 23 | 20 or 21 or 22 | 211820 |
| 24 | diagnosis/ | 1066933 |
| 25 | "sensitivity and specificity"/ | 231631 |
| 26 | 9 or 10 or 11 or 24 or 25 | 1642613 |
| 27 | brief cognitive tests.ti,ab. | 66 |
| 28 | cognitive screen*.ti,ab. | 1402 |
| 29 | ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. | 222 |
| 30 | cog*.ti,ab. | 368545 |
| 31 | 27 or 28 or 29 or 30 | 368615 |
| 32 | "primary care".ti,ab. | 100669 |
| 33 | "general practic*".ti,ab. | 40969 |
| 34 | "GP".ti,ab. | 41972 |
| 35 | 32 or 33 or 34 | 169068 |
| 36 | 8 and 26 and 31 | 9009 |

N.B. systematic review search filter

- exp review/
- (literature adj3 review$).ti,ab.
- exp meta analysis/
- exp "Systematic Review"/
- or/1-4
- (medline or medlars or embase or pubmed or cinahl or amed or psychlit or psyclit or psychinfo or psycinfo or scisearch or cochrane).ti,ab.
- RETRACTED ARTICLE/
- 6 or 7

- 5 and 8
- (systematic$ adj2 (review$ or overview)).ti,ab.
- (meta?anal$ or meta anal$ or meta-anal$ or metaanal$ or metanal$).ti,ab.
- 9 or 10 or 11

From BMJ Clinical Evidence strategy [undated] [Ovid] was used.

### 4.7.1. Search strategy, formatted for OVID MEDLINE

Ovid MEDLINE(R) 1946 to August Week 2 2015

| | | |
|---|---|---|
| 1 | dementia/ or aids dementia complex/ or alzheimer disease/ or aphasia, primary progressive/ or creutzfeldt-jakob syndrome/ or dementia, vascular/ or diffuse neurofibrillary tangles with calcification/ or frontotemporal lobar degeneration/ or huntington disease/ or kluver-bucy syndrome/ or lewy body disease/ | 127632 |
| 2 | Dementia/ | 39509 |
| 3 | exp dementia/ | 129660 |
| 4 | exp memory disorder/ | 24008 |
| 5 | dementia.ti,ab. | 67035 |
| 6 | alzheimer*.ti,ab. | 92544 |
| 7 | 1 or 2 or 3 or 4 or 5 or 6 | 188618 |
| 8 | systematic*.ti,ab. | 221142 |
| 9 | meta-analysis.ti,ab. | 57552 |
| 10 | "systematic review".ti,ab. | 50417 |
| 11 | (((comprehensive* or integrative or systematic*) adj3 (bibliographic* or review* or literature)) or (meta-analy* or metaanaly* or "research synthesis" or ((information or data) adj3 synthesis) or (data adj2 extract*))).ti,ab. or (cinahl or (cochrane adj3 trial*) or embase or medline or psyclit or (psycinfo not "psycinfo database") or pubmed or scopus or "sociological abstracts" or "web of science").ab. or ("cochrane database of systematic reviews" or evidence report technology assessment or evidence report technology assessment summary).jn. or Evidence Report: Technology Assessment*.jn. or ((review adj5 (rationale or evidence)).ti,ab. and review.pt.) or meta-analysis as topic/ or Meta-Analysis.pt. | 212532 |
| 12 | diagnosis/ | 17054 |
| 13 | "sensitivity and specificity"/ | 295160 |
| 14 | 8 or 9 or 10 or 12 or 13 | 562806 |
| 15 | brief cognitive tests.ti,ab. | 41 |
| 16 | cognitive screen*.ti,ab. | 735 |
| 17 | ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. | 161 |
| 18 | cog*.ti,ab. | 243367 |
| 19 | 15 or 16 or 17 or 18 | 243415 |
| 20 | "primary care".ti,ab. | 71023 |
| 21 | "general practic*".ti,ab. | 32013 |
| 22 | "GP".ti,ab. | 27969 |
| 23 | 20 or 21 or 22 | 121226 |
| 24 | 7 and 14 and 19 | 2660 |

N.B. systematic review search filter

(((comprehensive* or integrative or systematic*) adj3 (bibliographic* or review* or literature)) or (meta-analy* or metaanaly* or "research synthesis" or ((information or data) adj3 synthesis) or (data adj2 extract*))).ti,ab. or (cinahl or (cochrane adj3 trial*) or embase or

medline or psyclit or (psycinfo not "psycinfo database") or pubmed or scopus or "sociological abstracts" or "web of science").ab. or ("cochrane database of systematic reviews" or evidence report technology assessment or evidence report technology assessment summary).jn. or Evidence Report: Technology Assessment*.jn. or ((review adj5 (rationale or evidence)).ti,ab. and review.pt.) or meta-analysis as topic/ or Meta-Analysis.pt.

From University of Texas School of Public Health (Search filters for systematic reviews and meta-analyses. Accessed 06 Dec 2013), was used.

### 4.7.2. Search strategy, formatted for Psychinfo (OVID)

1806 to August Week 2 2015

| | | |
|---|---|---|
| 1 | dementia/ or aids dementia complex/ or alzheimer disease/ or aphasia, primary progressive/ or creutzfeldt-jakob syndrome/ or dementia, vascular/ or diffuse neurofibrillary tangles with calcification/ or frontotemporal lobar degeneration/ or huntington disease/ or kluver-bucy syndrome/ or lewy body disease/ | 55725 |
| 2 | exp dementia/ | 57951 |
| 3 | exp Neuropsychological Assessment/ | 14779 |
| 4 | exp Cognitive Assessment/ | 3559 |
| 5 | dementia.ti,ab. | 46844 |
| 6 | alzheimer*.ti,ab. | 44055 |
| 7 | 1 or 2 or 3 or 4 or 5 or 6 | 90002 |
| 8 | systematic*.ti,ab. | 85202 |
| 9 | meta-analysis.ti,ab. | 16369 |
| 10 | "systematic review".ti,ab. | 12154 |
| 11 | (((comprehensive* or integrative or systematic*) adj3 (bibliographic* or review* or literature)) or (meta-analy* or metaanaly* or "research synthesis" or ((information or data) adj3 synthesis) or (data adj2 extract*))).ti,ab,id. or ((review adj5 (rational or evidence)).ti,ab,id. and "Literature Review".md.) or (cinahl or (cochrane adj3 trial*) or embase or medline or psyclit or pubmed or scopus or "sociological abstracts" or "web of science").ab. or ("systematic review" or "meta analysis").md. | 54964 |
| 12 | 8 or 9 or 10 or 11 | 121571 |
| 13 | diagnosis/ | 36489 |
| 14 | brief cognitive tests.ti,ab. | 34 |
| 15 | cognitive screen*.ti,ab. | 708 |
| 16 | ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. | 132 |
| 17 | cog*.ti,ab. | 347512 |
| 18 | 14 or 15 or 16 or 17 | 347557 |
| 19 | 7 and 12 and 13 and 18 | 102 |

N.B. Search filters from University of Texas School of Public Health. Search filters for systematic reviews and meta-analyses. Accessed 06 Dec 2013. [Ovid] incorporated as they added ~30 hits.

### 4.7.3. Appendix 4 Cochrane reviews assessed for this overview

| Title | Include/ exclude |
|---|---|
| Arevalo-Rodriguez I, Smailagic N, Roqué i Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, Pedraza OL, Bonfill Cosp X, Cullum S. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database of Systematic Reviews 2015, Issue 3. Art. No.: CD010783. DOI: 10.1002/14651858.CD010783.pub2. | Include |
| Chan CCH, Fage BA, Smailagic N, Gill SS, Herrmann N, Nikolaou V, Seitz DP. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a secondary care setting (Protocol). Cochrane Database of Systematic Reviews 2014, Issue 12. Art. No.: CD011414. DOI: 10.1002/14651858.CD011414. | Exclude |
| Creavin ST, Wisniewski S, Noel-Storr AH, Trevelyan CM, Hampton T, Rayment D, Thom VM, Nash KJE, Elhamoui H, Milligan R, Patel AS, Tsivos DV, Wing T, Phillips E, Kellman SM, Shackleton HL, Singleton GF, Neale BE, Watton ME, Cullum S. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. Cochrane Database of Systematic Reviews 2016, Issue 1. Art. No.: CD011145. DOI: 10.1002/14651858.CD011145.pub2. | Include |
| Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. Cochrane Database of Systematic Reviews 2015, Issue 10. Art. No.: CD010775. DOI: 10.1002/14651858.CD010775.pub2. | Include |
| Fage BA, Chan CCH, Gill SS, Noel-Storr AH, Herrmann N, Smailagic N, Nikolaou V, Seitz DP. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting. Cochrane Database of Systematic Reviews 2015, Issue 2. Art. No.: CD010860. DOI: 10.1002/14651858.CD010860.pub2.. | Exclude |
| Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. Cochrane Database of Systematic Reviews 2015, Issue 3. Art. No.: CD010772. DOI: 10.1002/14651858.CD010772.pub2. | Exclude |
| Harrison, J. K., P. Fearon, A. H. Noel-Storr, R. McShane, D. J. Stott and T. J. Quinn (2014) Informant Questionnaire on Cognitive Decline in the Elderly IQCODE for the diagnosis of dementia within a general practice (primary care) setting. The Cochrane database of systematic reviews 7: CD010771.Cochrane Database of Systematic Reviews DOI: 10.1002/14651858.CD010783.pub2. | Include |
| Hendry, K., A. Lees Rosalind, R. McShane, H. Noel-Storr Anna, J. Stott David and J. Quinn Terry (2014)AD-8 for diagnosis of dementia across a variety of healthcare settings. Cochrane Database of Systematic Reviews DOI: 10.1002/14651858.cd011121. | Include |
| Lees Rosalind, A., J. Stott David, R. McShane, H. Noel-Storr Anna and J. Quinn Terry (2014) Informant Questionnaire on Cognitive Decline in the Elderly IQCODE for the early diagnosis of dementia across a variety of healthcare settings. Cochrane Database of Systematic Reviews DOI: 10.1002/14651858.cd011333. | Include |
| Quinn Terry, J., P. Fearon, H. Noel-Storr Anna, C. Young, R. McShane and J. Stott David (2014) "Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations." Cochrane Database of Systematic Reviews DOI: 10.1002/14651858.CD010079.pub2. | Exclude |

| | |
|---|---|
| Seitz Dallas, P., A. Fage Bruce, C. H. Chan Calvin, S. Gill Sudeep, N. Herrmann, N. Smailagic and V. Nikolaou (2014) Mini-Cog for the diagnosis of Alzheimer?s disease dementia and other dementias within a primary care setting. Cochrane Database of Systematic Reviews DOI: 10.1002/14651858.cd011415. | Include |

### 4.7.4. Appendix 5 Exclusions at full text screening

| | Full reference | Reason |
|---|---|---|
| 1 | Al-Qazzaz, N. A., S. H. Ali, S. A. Ahmad and S. Islam (2014). "Cognitive assessments for the early diagnosis of dementia after stroke." Neuropsychiatric Disease and Treatment **10**: 1743-1751. | Design |
| 2 | Arevalo-Rodriguez, I., O. Segura, I. Sola, X. Bonfill, E. Sanchez and P. Alonso-Coello (2014). "Diagnostic tools for alzheimer's disease dementia and other dementias: An overview of diagnostic test accuracy (DTA) systematic reviews." BMC Neurology **14**(1). | Design |
| 3 | Borson, S., M. Brush, E. Gil, J. Scanlan, P. Vitaliano, J. Chen, J. Cashman, M. M. Sta Maria, R. Barnhart and J. Roques "The Clock Drawing Test: utility for dementia detection in multiethnic elders." Journals of Gerontology Series A-Biological Sciences & Medical Sciences **54**(11): M534-540. | Design |
| 4 | Canning, S. J., L. Leach, D. Stuss, L. Ngo and S. E. Black "Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia." Neurology **62**(4): 556-562. | Design |
| 5 | Chan, C. C. and P. T. Lam (2005). "Update on dementia - Part 1: Mild cognitive impairment, screening and diagnostic assessment." Hong Kong Practitioner **27**(6): 235-241. | Design |
| 6 | Cossa, F. M., S. Della Sala, M. Musicco, H. Spinnler and M. C. Ubezio "Comparison of two scoring systems of the Mini-Mental State Examination as a screening test for dementia." Journal of Clinical Epidemiology **50**(8): 961-965. | Design |
| 7 | Crawford, S., J. Evans, L. Whitnall and J. A. Robertson (2011). "A systematic review of the accuracy and clinical utility of the addenbrooke's cognitive examination and the addenbrooke's cognitive examination - Revised in the diagnosis of dementia." Brain Impairment **12**: 4. | Setting |
| 8 | Darzins, P. and D. LoGiudice (1999). "Clinical testing in general practice. What is the evidence?" Australian family physician **28**(12): 1241-1244. | Design |
| 9 | Diesfeldt, H. F. "[Discrepancies between the IQCODE (Informant Questionnaire on Cognitive Decline in the Elderly) and cognitive test performance]." Tijdschrift voor Gerontologie en Geriatrie **38**(5): 225-236. | Design |
| 10 | Evans, I. E. M., E. Kuzma, I. A. Lang, A. L. R. Adlam and D. J. Llewellyn (2014). "Which brief assessment measures for dementia are currently recommended for use in primary care? a systematic review." Alzheimer's and Dementia **10**: P439. | Design |
| 11 | Fage, B. A., C. C. Chan, S. S. Gill, A. H. Noel-Storr, N. Herrmann, N. Smailagic, V. Nikolaou and D. P. Seitz (2015). "Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting." The Cochrane database of systematic reviews **2**: CD010860. | Setting |
| 12 | Flicker, L., D. Logiudice, J. B. Carlin and D. Ames "The predictive value of dementia screening instruments in clinical populations." International Journal of Geriatric Psychiatry **12**(2): 203-209. | Design |
| 13 | Fong, T. G., M. A. Fearing, R. N. Jones, P. Shi, E. R. Marcantonio, J. L. Rudolph, F. M. Yang, D. K. Kiely and S. K. Inouye "Telephone interview for cognitive status: Creating a crosswalk with the Mini-Mental State Examination." Alzheimer's & Dementia **5**(6): 492-497. | Design |
| 14 | Fuzikawa, C. S., E. Uchoa and M. F. Lima-Costa (2003). "Clock drawing test: A review on this cognitive screening test. [Portuguese]." Jornal Brasileiro de Psiquiatria **52**(3): 223-235. | Setting |
| 15 | Gainotti, G., C. Marra, G. Villa, V. Parlato and F. Chiarotti "Sensitivity and specificity of some neuropsychological markers of Alzheimer dementia." Alzheimer Disease & Associated Disorders **12**(3): 152-162. | Design |
| | Harvan, J.R. and Cotter, V.T., 2006. An evaluation of dementia screening in the primary care setting. *Journal of the American Academy of Nurse Practitioners*, *18*(8), pp.351-360 | Design |
| 16 | Holsinger, T., B. L. Plassman and K. M. Stechuchak (2012). "The Mini-Cog had sensitivity similar to the longer 3MS for detecting cognitive impairment or dementia." Annals of Internal Medicine **157**(8): JC4-JC8. | Design |

| | | |
|---|---|---|
| 17 | Iliffe, S., L. Robinson, C. Brayne, C. Goodman, G. Rait, J. Manthorpe, P. Ashley and N. P. C. C. S. G. De "Primary care and dementia: 1. diagnosis, screening and disclosure." International Journal of Geriatric Psychiatry **24**(9): 895-901. | Design |
| 18 | Inouye, S. K., J. T. Robison, T. E. Froehlich and E. D. Richardson "The time and change test: a simple screening test for dementia." Journals of Gerontology Series A-Biological Sciences & Medical Sciences **53**(4): M281-286. | Design |
| 19 | Ismail, Z., T. K. Rajji and K. I. Shulman (2010). "Brief cognitive screening instruments: an update." Int J Geriatr Psychiatry **25**(2): 111-120. | Design |
| | Jacova, C., Kertesz, A., Blair, M., Fisk, J.D. and Feldman, H.H., 2007. Neuropsychological testing and assessment for dementia. *Alzheimer's & Dementia*, *3*(4), pp.299-317. | Setting |
| 20 | Jacqmin-Gadda, H., C. Fabrigoule, D. Commenges, L. Letenneur and J. F. Dartigues "A cognitive screening battery for dementia in the elderly." Journal of Clinical Epidemiology **53**(10): 980-987. | Design |
| 21 | Jitapunkul, S., I. Pillay and S. Ebrahim "The abbreviated mental test: its use and validity." Age & Ageing **20**(5): 332-336. | Design |
| 22 | Jorm, A. F. (1997). "Methods of screening for dementia: A meta-analysis of studies comparing an informant questionnaire with a brief cognitive test." Alzheimer Disease and Associated Disorders **11**(3): 158-162. | Setting |
| 23 | Kalbe, E., J. Kessler, P. Calabrese, R. Smith, A. P. Passmore, M. Brand and R. Bullock "DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia." International Journal of Geriatric Psychiatry **19**(2): 136-143. | Design |
| 24 | Larner A., M., A. J. (2014). "A meta-analysis of the accuracy of the Addenbrooke's Cognitive Examination (ACE) and the Addenbrooke's Cognitive Examination-Revised (ACE-R) in the detection of dementia." International psychogeriatrics / IPA **26**(4): 555-563. | Setting |
| 25 | Larner, A. and A. J. Mitchell (2014). "Ace and ACE-R for diagnosis of dementia: A meta-analysis." Journal of Neurology, Neurosurgery and Psychiatry **85 (10)**: A16. | Setting |
| 26 | Lorentz, W. J., J. M. Scanlan and S. Borson (2002). "Brief screening tests for dementia." Can J Psychiatry **47**(8): 723-733. | Design |
| 27 | Mahoney, R., K. Johnston, C. Katona, K. Maxmin and G. Livingston "The TE4D-Cog: a new test for detecting early dementia in English-speaking populations." International Journal of Geriatric Psychiatry **20**(12): 1172-1179. | Design |
| 28 | Milne, A., A. Culverwell, R. Guss, J. Tuppen and R. Whelton (2008). "Screening for dementia in primary care: A review of the use, efficacy and quality of measures." International Psychogeriatrics **20**(5): 911-926. | Design |
| 29 | Mundt, J. C., K. L. Ferber, M. Rizzo and J. H. Greist "Computer-automated dementia screening using a touch-tone telephone." Archives of Internal Medicine **161**(20): 2481-2487. | Design |
| 30 | Quinn Terry, J., P. Fearon, H. Noel-Storr Anna, C. Young, R. McShane and J. Stott David (2014) "Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations." Cochrane Database of Systematic Reviews DOI: 10.1002/14651858.CD010079.pub2. | Setting |
| 31 | Rait, G., M. Morley, A. Burns, R. Baldwin, C. Chew-Graham and A. S. St Leger "Screening for cognitive impairment in older African-Caribbeans." Psychological Medicine **30**(4): 957-963. | Design |
| 32 | Rous, R. S., C. R. Housden, L. M. Lewis, A. Filby, M. J. Taylor, A. D. Blackwell and J. H. Barnett (2014). "The sensitivity and specificity of computerised or paper-and-pencil cognitive assessments used in primary care impact the cost-effectiveness of the dementia diagnostic pathway." Alzheimer's and Dementia **10**: P566. | Design |
| 33 | Stolwyk, R. J., M. H. O'Neill, A. J. D. McKay and D. K. Wong (2014). "Are cognitive screening tools sensitive and specific enough for use after stroke?: A systematic literature review." Stroke **45**(10): 3129-3134. | Setting |
| 34 | Stuss, D. T., N. Meiran, D. A. Guzman, G. Lafleche and J. Willmer "Do long tests yield a more accurate diagnosis of dementia than short tests? A comparison of 5 neuropsychological tests." Archives of Neurology **53**(10): 1033-1039. | Design |

| | | |
|---|---|---|
| 35 | Tang, W. K., S. S. Chan, H. F. Chiu, K. S. Wong, T. C. Kwok, V. Mok and G. S. Ungvari "Can IQCODE detect poststroke dementia?" International Journal of Geriatric Psychiatry **18**(8): 706-710. | Design |
| 36 | Tangalos, E. G., G. E. Smith, R. J. Ivnik, R. C. Petersen, E. Kokmen, L. T. Kurland, K. P. Offord and J. E. Parisi "The Mini-Mental State Examination in general medical practice: clinical utility and acceptance." Mayo Clinic Proceedings **71**(9): 829-837. | Design |
| 37 | Tierney, M. C. and M. A. Lermer (2010). "Computerized cognitive assessment in primary care to identify patients with suspected cognitive impairment." Journal of Alzheimer's Disease **20**(3): 823-832. | Condition |
| 38 | Uhlmann, R. F. and E. B. Larson "Effect of education on the mini-mental state examination as a screening test for dementia." Journal of the American Geriatrics Society **39**(9): 876-880. | Design |
| 39 | Uhlmann, R. F., T. S. Rees, B. M. Psaty and L. G. Duckert "Validity and reliability of auditory screening tests in demented and non-demented older adults." Journal of General Internal Medicine **4**(2): 90-96. | Design |
| 40 | Valverde, A. H., A. Jimenez-Escrig, J. Gobernado and M. Baron "A short neuropsychologic and cognitive evaluation of frontotemporal dementia." Clinical Neurology & Neurosurgery **111**(3): 251-255. | Design |
| 41 | van Gorp, W. G., T. D. Marcotte, D. Sultzer, C. Hinkin, M. Mahler and J. L. Cummings "Screening for dementia: comparison of three commonly used instruments." Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society **21**(1): 29-38. | Design |
| 42 | Watson, Y. I., C. L. Arfken and S. J. Birge "Clock completion: an objective screening test for dementia." Journal of the American Geriatrics Society **41**(11): 1235-1240. | Design |
| 43 | Wind, A. W., F. G. Schellevis, G. Van Staveren, R. P. Scholten, C. Jonker and J. T. Van Eijk "Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice." International Journal of Geriatric Psychiatry **12**(1): 101-108. | Design |

## Appendix 6 AMSTAR checklist [212]

**AMSTAR – a measurement tool to assess the methodological quality of systematic reviews.**

**1. Was an 'a priori' design provided?**
The research question and inclusion criteria should be established before the conduct of the review.

*Note: Need to refer to a protocol, ethics approval, or pre-determined/a priori published research objectives to score a "yes."*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**2. Was there duplicate study selection and data extraction?**
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

*Note: 2 people do study selection, 2 people do data extraction, consensus process or one person checks the other's work.*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**3. Was a comprehensive literature search performed?**
At least two electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

*Note: If at least 2 sources + one supplementary strategy used, select "yes" (Cochrane register/Central counts as 2 sources; a grey literature search counts as supplementary).*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?**
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

*Note: If review indicates that there was a search for "grey literature" or "unpublished literature," indicate "yes." SIGLE database, dissertations, conference proceedings, and trial registries are all considered grey for this purpose. If searching a source that contains both grey and non-grey, must specify that they were searching for grey/unpublished lit.*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**5. Was a list of studies (included and excluded) provided?**
A list of included and excluded studies should be provided.

*Note: Acceptable if the excluded studies are referenced. If there is an electronic link to the list but the link is dead, select "no."*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**6. Were the characteristics of the included studies provided?**
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

*Note: Acceptable if not in table format as long as they are described as above.*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**7. Was the scientific quality of the included studies assessed and documented?**
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

*Note: Can include use of a quality scoring tool or checklist, e.g., Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for EACH study ("low" or "high" is fine, as long as it is clear which studies scored "low" and which scored "high"; a summary score/range for all studies is not acceptable).*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**8. Was the scientific quality of the included studies used appropriately in formulating conclusions?**
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

*Note: Might say something such as "the results should be interpreted with caution due to poor quality of included studies." Cannot score "yes" for this question if scored "no" for question 7.*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**9. Were the methods used to combine the findings of studies appropriate?**
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., Chi-squared test for homogeneity, $I^2$). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?).

*Note: Indicate "yes" if they mention or describe heterogeneity, i.e., if they explain that they cannot pool because of heterogeneity/variability between interventions.*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**10. Was the likelihood of publication bias assessed?**
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test, Hedges-Olken).

*Note: If no test values or funnel plot included, score "no". Score "yes" if mentions that publication bias could not be assessed because there were fewer than 10 included studies.*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

**11. Was the conflict of interest included?**
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

*Note: To get a "yes," must indicate source of funding or support for the systematic review AND for each of the included studies.*

□ Yes
□ No
□ Can't answer
□ Not applicable

---

Shea *et al. BMC Medical Research Methodology* 2007 **7**:10 doi:10.1186/1471-2288-7-10

*Additional notes (in italics) made by Michelle Weir, Julia Worswick, and Carolyn Wayne based on conversations with Bev Shea and/or Jeremy Grimshaw in June and October 2008 and July and September 2010.*

## Appendix 7 ROBIS checklist [213] [tailored by Harriet Hunt]

### ROBIS: Tool to assess risk of bias in systematic reviews

### Phase 1: Assessing relevance (Optional)

ROBIS is designed to assess the risk of bias in reviews with questions relating to interventions, aetiology, diagnosis and prognosis. State your overview/guideline question (target question) and the question being addressed in the review being assessed:

**Intervention reviews:**

| Category | Target question (e.g. overview or guideline) | Review being assessed |
|---|---|---|
| Patients/Population(s): | | |
| Intervention(s): | | |
| Comparator(s): | | |
| Outcome(s): | | |

**For aetiology reviews:**

| Category | Target question (e.g. overview or guideline) | Review being assessed |
|---|---|---|
| Patients/Population(s): | | |
| Exposure(s) and comparator(s): | | |
| Outcome(s): | | |

**For DTA reviews:**

| Category | Target question (e.g. overview or guideline) | Review being assessed |
|---|---|---|
| Patients): | Primary care patients with possible dementia | |
| Index test(s): | Brief cognitive assessments (all types) | |
| Reference standard: | Clinical diagnosis and follow-up according to recognised diagnostic criteria e.g. DSM,ICD and CAMDEX | |
| Target condition: | Probable all-cause dementia | |

**For prognostic reviews:**

| Category | Target question (e.g. overview or guideline) | Review being assessed |
|---|---|---|
| Patients: | | |
| Outcome to be predicted: | | |
| Intended use of model: | | |
| Intended moment in time: | | |

| Does the question addressed by the review match the target question? | YES/NO/UNCLEAR |
|---|---|

**Phase 2: Identifying concerns with the review process**

| **DOMAIN 1: STUDY ELIGIBILITY CRITERIA** | |
|---|---|
| Describe the study eligibility criteria, any restrictions on eligibility and whether there was evidence that objectives and eligibility criteria were pre-specified: | |
| 1.1 Did the review adhere to pre-defined objectives and eligibility criteria? | Y/PY/PN/N/NI |
| 1.2 Were the eligibility criteria appropriate for the review question? | Y/PY/PN/N/NI |
| 1.3 Were eligibility criteria unambiguous? | Y/PY/PN/N/NI |
| 1.4 Were all restrictions in eligibility criteria based on study characteristics appropriate (e.g. date, sample size, study quality, outcomes measured)? | Y/PY/PN/N/NI |
| 1.5 Were any restrictions in eligibility criteria based on sources of information appropriate (e.g. publication status or format, language, availability of data)? | Y/PY/PN/N/NI |
| Concerns regarding specification of study eligibility criteria | LOW/HIGH/UNCLEAR |
| Rationale for concern: | |

| **DOMAIN 2: IDENTIFICATION AND SELECTION OF STUDIES** | |
|---|---|
| Describe methods of study identification and selection (e.g. number of reviewers involved): | |
| 2.1 Did the search include an appropriate range of databases/electronic sources for published and unpublished reports? | Y/PY/PN/N/NI |
| 2.2 Were methods additional to database searching used to identify relevant reports? | Y/PY/PN/N/NI |
| 2.3 Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible? | Y/PY/PN/N/NI |
| 2.4 Were restrictions based on date, publication format, or language appropriate? | Y/PY/PN/N/NI |
| 2.5 Were efforts made to minimise error in selection of studies? | Y/PY/PN/N/NI |
| Concerns regarding methods used to identify and/or select studies | LOW/HIGH/UNCLEAR |
| Rationale for concern: | |

| **DOMAIN 3: DATA COLLECTION AND STUDY APPRAISAL** | |
|---|---|
| Describe methods of data collection, what data were extracted from studies or collected through other means, how risk of bias was assessed (e.g. number of reviewers involved) and the tool used to assess risk of bias: | |
| 3.1 Were efforts made to minimise error in data collection? | Y/PY/PN/N/NI |
| 3.2 Were sufficient study characteristics available for both review authors and readers to be able to interpret the results? | Y/PY/PN/N/NI |
| 3.3 Were all relevant study results collected for use in the synthesis? | Y/PY/PN/N/NI |
| 3.4 Was risk of bias (or methodological quality) formally assessed using appropriate criteria? | Y/PY/PN/N/NI |
| 3.5 Were efforts made to minimise error in risk of bias assessment? | Y/PY/PN/N/NI |
| Concerns regarding methods used to collect data and appraise studies | LOW/HIGH/UNCLEAR |
| Rationale for concern: | |

| DOMAIN 4: SYNTHESIS AND FINDINGS | |
|---|---|
| Describe synthesis methods: | |
| 4.1 Did the synthesis include all studies that it should? | Y/PY/PN/N/NI |
| 4.2 Were all pre-defined analyses reported or departures explained? | Y/PY/PN/N/NI |
| 4.3 Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies? | Y/PY/PN/N/NI |
| 4.4 Was between-study variation (heterogeneity) minimal or addressed in the synthesis? | Y/PY/PN/N/NI |
| 4.5 Were the findings robust, e.g. as demonstrated through funnel plot or sensitivity analyses? | Y/PY/PN/N/NI |
| 4.6 Were biases in primary studies minimal or addressed in the synthesis? | Y/PY/PN/N/NI |
| Concerns regarding the synthesis and findings<br>Rationale for concern: | LOW/HIGH/UNCLEAR |

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

### Phase 3: Judging risk of bias

Summarize the concerns identified during the Phase 2 assessment:

| Domain | Concern | Rationale for concern |
|---|---|---|
| 1. Concerns regarding specification of study eligibility criteria | | |
| 2. Concerns regarding methods used to identify and/or select studies | | |
| 3. Concerns regarding used to collect data and appraise studies | | |
| 4. Concerns regarding the synthesis and findings | | |

| RISK OF BIAS IN THE REVIEW | |
|---|---|
| Describe whether conclusions were supported by the evidence: | |
| A. Did the interpretation of findings address all of the concerns identified in Domains 1 to 4? | Y/PY/PN/N/NI |
| B. Was the relevance of identified studies to the review's research question appropriately considered? | Y/PY/PN/N/NI |
| C. Did the reviewers avoid emphasizing results on the basis of their statistical significance? | Y/PY/PN/N/NI |
| Risk of bias in the review | RISK: LOW/HIGH/UNCLEAR |

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

| Glossary | |
|---|---|
| **ADI** | Alzheimer's disease International |
| **AGECAT** | Automated Geriatric Examination for Computer Assisted Taxonomy |
| **AMT(S)** | Abbreviated Mental Test Score |
| **CAMCOG** | The Cambridge Cognitive Examination |
| **CAMDEX** | Cambridge Mental Disorders of the Elderly Examination |
| **CASI** | Cognitive Abilities Screening Instrument |
| **CCCDTD** | Canadian Consensus Conferences on the Diagnosis and Treatment of Dementia |
| **CCG** | Clinical Commissioning Group |
| **CDPC** | Cognitive Decline Partnership Centre |
| **CDR** | Clinical Dementia Rating |
| **CDT** | clock drawing test |
| **CERAD** | Consortium to Establish a Registry for Alzheimer's Disease |
| **CPG** | Clinical Practice Guideline |
| **DSM-III/ III-R/ IV/ IV-R** | The Diagnostic and Statistical Manual of Mental Disorders (version 3/ version 3 revised/ version 4/ version 4 revised |
| **EMBASE** | Excerpta Medica dataBASE |
| **FAB** | Frontal Assessment Battery |
| **FAQ** | Functional Activities Questionnaire |
| **FCSRT** | Free and Cued Selective Reminding Test |
| **GMS-AGECAT** | Geriatric Mental State Schedule - Automated Geriatric Examination for Computer Assisted Taxonomy |
| **GP** | General Practitioner |
| **GPCOG** | The General Practitioner assessment of Cognition |
| **ICD-10** | International Classification of Disease – version 10 |
| **IQCODE** | Informant Questionnaire on Cognitive Decline in the Elderly |
| **IPA/WHO criteria** | International Psychogeriatric Association/World Health organisation criteria |
| **KICA-Cog** | Kimberley Indigenous Cognitive Assessment |
| **KICA-Screen** | Kimberley Indigenous Cognitive screening tool |
| **MCI** | Mild Cognitive Impairment |
| **MEDLINE** | Medical Literature Analysis and Retrieval System Online |
| **MeSH** | Medical Subject Headings |

| | |
|---|---|
| **MIS** | Memory Impairment Screen |
| **MMSE** | Mini mental state examination |
| **MoCA** | Montreal Cognitive Assessment |
| **NHS** | National Health Service (UK) |
| **NICE** | National Institute for Health and Care Excellence (UK) |
| **NINDCS-ADRDA** | National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association |
| **NINCDS-AIREN** | National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l'Enseignement en Neurosciences |
| **NINCDR-CERAD** | National Institute of Neurological Disorders - Clinical Dementia Rating -Consortium to Establish a Registry for Alzheimer's Disease |
| **PAS** | Psychogeriatric Assessment Scale |
| **PCL** | Prueba cognitive de leganes |
| **PROSPERO** | International Prospective Register of Systematic Reviews |
| **PsychInfo** | Database of abstracts of literature in the field of psychology. |
| **RUDAS** | The Rowland Universal Dementia Assessment Scale |
| **SASSI** | Short and Sweet Screening Instrument |
| **SIS** | Six item screener |
| **SPMSQ** | Short portable mental status questionnaire |
| **TICS** | Telephone Interview for Cognitive Status |
| **TRIP** | Turning Research Into Practice database |
| **TYM** | Test Your Memory |
| **VF-an** | Verbal Fluency - animals |
| **WHO** | World Health Organisation |

# 5. How accurate are GPCOG and MMSE in identifying dementia when directly compared to each other?

*Any truth is better than indefinite doubt.*

Conan Doyle, A (1894) The memoirs of Sherlock Holmes. London, George Newnes

The overview described in Chapter 3 revealed that whilst a large volume of systematic review evidence exists, the majority of evidence is often derived from a comparison of diagnostic accuracy across different studies (termed "indirect comparisons"). When diagnostic accuracy is compared across different studies, it is challenging to account for all the differences in setting, population, test administration and other factors which affect performance, yet these variations may all have a bearing on how the test is used and performs in a clinical setting. In order to minimise the potential influence of extraneous variables between different studies, a systematic review was conducted which exclusively assessed direct comparisons of MMSE and GPCOG, where both tests of interest had been compared within the same study. It was hypothesised that such a systematic review of direct comparisons would demonstrate which BCA had superior diagnostic accuracy within the same study population. In reality, this systematic review revealed a complex set of problems around both the findings of included studies, and methodological challenges of conducting a systematic review of direct comparisons.

The difficulties inherent in this approach to evaluating diagnostic accuracy are discussed, and potential considered for combining the most useful information to form a more complete picture of benefit for clinical decision making. These findings lead on to Chapter 6 in which these factors are considered within the context of clinical practice using survey research, and Chapter 7 where all findings are discussed in the broader context of the entire thesis.

## 5.1. Background to the review

Many systematic reviews[102,104-106,279-281] have explored the individual diagnostic accuracy of brief cognitive assessments for cognitive impairment as part of the process for identifying dementia in isolation, and across a range of populations and settings including primary care, community and memory clinics. Why it is then

necessary to conduct yet more syntheses of evidence when the evidence clearly already exists? There are several reasons. Firstly, many systematic reviews of diagnostic accuracy compare different test performance across different studies. This means that observed differences between tests and test accuracy may not be down to differing performance between various brief cognitive assessments, but may be due to other confounding factors such as variation in test threshold used, or differences in testing populations.

The overview of diagnostic accuracy reported in Chapter 3 revealed many problems with the summarised evidence on the accuracy of brief cognitive assessments. One specific challenge is that the focus has been on the the accuracy of individual brief cognitive assessments, rather than the degree to which one brief cognitive assessment compares to another. Within the overview, eight[99,105,120,141,148,222,231,236] of the 13 included systematic reviews reported data from indirect comparison studies (i.e. where data from individual studies assessing single brief cognitive assessments was compared within a single systematic review) and only one systematic review[125] reported data from a single direct within-study[142] comparison of MMSE, GPCOG and the Abbreviated Mental Test (AMT). Within the evidence making indirect comparisons, there appears to be under-recognition of the degree to which confounding may be responsible for variation between test accuracy results for one test relative to another. In addition, the measures which would usually be taken to reduce this variation within standard multivariate analyses (such as ensuring consistency in population characteristics, reference standards, and study designs[282]) do not appear to have been routinely applied in many accuracy studies.

Direct comparisons of test performance within the same study population are not often conducted, and this lack of directly-comparable data was the reason for carrying out this systematic review solely focussed upon direct comparisons of two brief cognitive assessments. In order to focus the question in this way, the results of the overview and advice from clinical colleagues were combined to identify the BCA comparison of most clinically relevance.

Two brief cognitive assessments were identified as suitable to compare against one another in order to assess diagnostic accuracy. These tests, the Mini Mental State Examination (MMSE) and General Practitioner Assessment of Cognition (GPCOG), were the two most frequently-assessed brief cognitive assessments in the 13

systematic reviews[88,103,105,106,120,125,141,148,221,222,231,233,236] included within the overview, with the MMSE featuring in eight reviews, and the GPCOG featuring in four reviews. The clock drawing test (CDT) was the third most frequently-assessed tool, also featuring in four reviews. The CDT was judged to be less comparable to the MMSE in terms of administration complexity, timing and domains assessed, relative to the GPCOG.

As the most frequently-assessed test within the overview, the MMSE is included as one of the index tests within this review as, whilst copyright restrictions are now enforced, it remains one of the most popular brief cognitive assessments employed in practice[88,116]. The MMSE is based on a 30 point scale of 11 questions testing five domains of cognitive function (orientation, registration, attention and calculation, recall and language)[115].

The GPCOG was the second most frequently-assessed index test within the overview featured in Chapter 3. The GPCOG is a publicly-available test in two sections: a patient examination (GPCOG-Patient) with a maximum score of nine (optimum performance) covering time orientation, clock drawing, reporting recent events and a word-recall task, and an optional informant questionnaire (GPCOG-Informant) with a maximum score of six with questions assessing the patient's memory of recent events and their executive function[142]. These details are summarised in Table 14. Within the survey of UK general practitioners (GPs) reported in Chapter 7 exploring how GPs choose and use brief cognitive assessments as part of the process for identifying dementia in primary care, MMSE and GPCOG were the two most frequently used assessments with 32% of respondents selecting each test. Finally, both tests have been developed independently whereas many other brief cognitive assessments share elements of the MMSE. MMSE was created by Folstein and colleagues[115] in 1975 with a highly selected group of 69 psychiatric inpatients representing a spectrum of different clinical conditions. GPCOG was developed in 2002 by Brodaty and colleagues[142] within a general practice population using a group of 283 community-dwelling participants, either with memory complaints of between 50-75 years old, or asymptomatic if over 75 years.

Both MMSE and GPCOG measure short term memory and visuospatial/constructional praxis, with minimal assessment data available on these domains for MMSE and moderate assessment of GPCOG[125] – this pattern is shown in Table 14. Of the tools

shown, only GPCOG measures semantic memory and frontal or executive function, with minimal assessment data available. Both tools measure orientation, with relatively extensive assessment information available for the MMSE and minimal assessment information for GPCOG. Attention/calculation is covered by both tools, with moderate assessment data on this domain available for MMSE and minimal assessment data available on GPCOG. Finally, language is covered by MMSE with moderate assessment data available, whereas there is no evidence that this has been specifically tested in GPCOG[125].

Table 14 Features of top three brief cognitive assessments identified within the overview

| Cognitive domains & other details | MMSE | GPCOG | CDT |
|---|---|---|---|
| **Short term memory** | + | ++ | - |
| **Semantic memory** | - | + | + |
| **Visuospatial/ constructional praxis** | + | ++ | ++ |
| **Frontal/ executive function** | - | + | + |
| **Orientation** | | +++ | + |
| **Attention/ calculation** | ++ | + | + |
| **Language** | ++ | - | - |
| **Informant component** | No | Yes | No |
| **Maximum score** | 30 | 15 | 1 |
| **Number of SR inclusions in overview (# of study inclusions across SRs)** | 8 (22) | 4 (2) | 4 (2) |

MMSE, mini mental state questionnaire; GPCOG, General Practitioner Assessment of Cognition; CDT, clock drawing test; BCA, brief cognitive assessment; SR, systematic review; Max., maximum; **+**, minimal assessment; **++**, moderate assessment; **+++**, relatively extensive assessment. Adapted from Woodford and George 2007[125]

## 5.2. Methods

### 5.2.1. Systematic Review Protocol

This review was registered on the International Prospective Register of Systematic Reviews (PROSPERO) reference 42015022078[283], and conducted in line with a full pre-published study protocol[284]. The protocol was published within an open access peer-reviewed journal and is available in Appendix 8. A pragmatic search strategy was used, refining searches that built upon studies identified as part of the overview of systematic reviews of the diagnostic accuracy of brief cognitive assessments for identifying dementia in primary care (described in Chapter 4 of this thesis). The review was conducted as reported within the protocol.

*Departures from the published protocol*

There were two departures from the published protocol. One was regarding stratification of GPCOG. Within the protocol, it was stated that the most clinically-relevant measures of GPCOG Patient, GPCOG Total and GPCOG Two stage (as GPCOG Informant has not been recommended for use by itself within a clinical

setting[284]) would be used. The studies identified for inclusion within this systematic review reported four individual measures for GPCOG, using GPCOG Informant, GPCOG Patient, GPCOG Total and GPCOG Two stage so on this occasion the data guided the decision and was and stratified across these four GPCOG subtypes, rather than the pre-stated and arguably more clinically-relevant three subtypes.

The other departure from the protocol was in data extraction where it had been stated in the protocol that the data abstraction form would be piloted with two included studies. In the event only one study was used to pilot the form, as it was a simple process and necessary amendments to the form were obvious after only one trial run.

### 5.2.2. Outcomes

*Primary outcome*

The primary outcome of this review was the comparative accuracy of the two tests assessed via direct comparisons, that is the diagnostic accuracy of the two tests compared within the same population in a single study.

*Secondary outcome*

The secondary outcome of the review was to identify other common test-related factors identified by included studies, such as acceptability or administration time. Whilst beyond the primary focus of test accuracy, these other factors may contribute to the overall usefulness of the tests when applied in a general practice setting, and these were incorporated in the review findings to try and make the most useful research and clinical recommendations.

### 5.2.3. Search methods from the overview

In order to build the search database for the overview of systematic reviews of the diagnostic accuracy of brief cognitive assessments for identifying dementia in primary care, the Cochrane Database of Systematic Reviews, EMBASE, MEDLINE and PsychINFO were searched for systematic reviews from inception until August 2015. Full search strategies are shown in Chapter 3, Appendices 3 & 4. The full search methods from the overview of systematic reviews summarising the accuracy of brief cognitive assessments for identifying dementia in primary care are presented in Chapter 3, section 3.3.

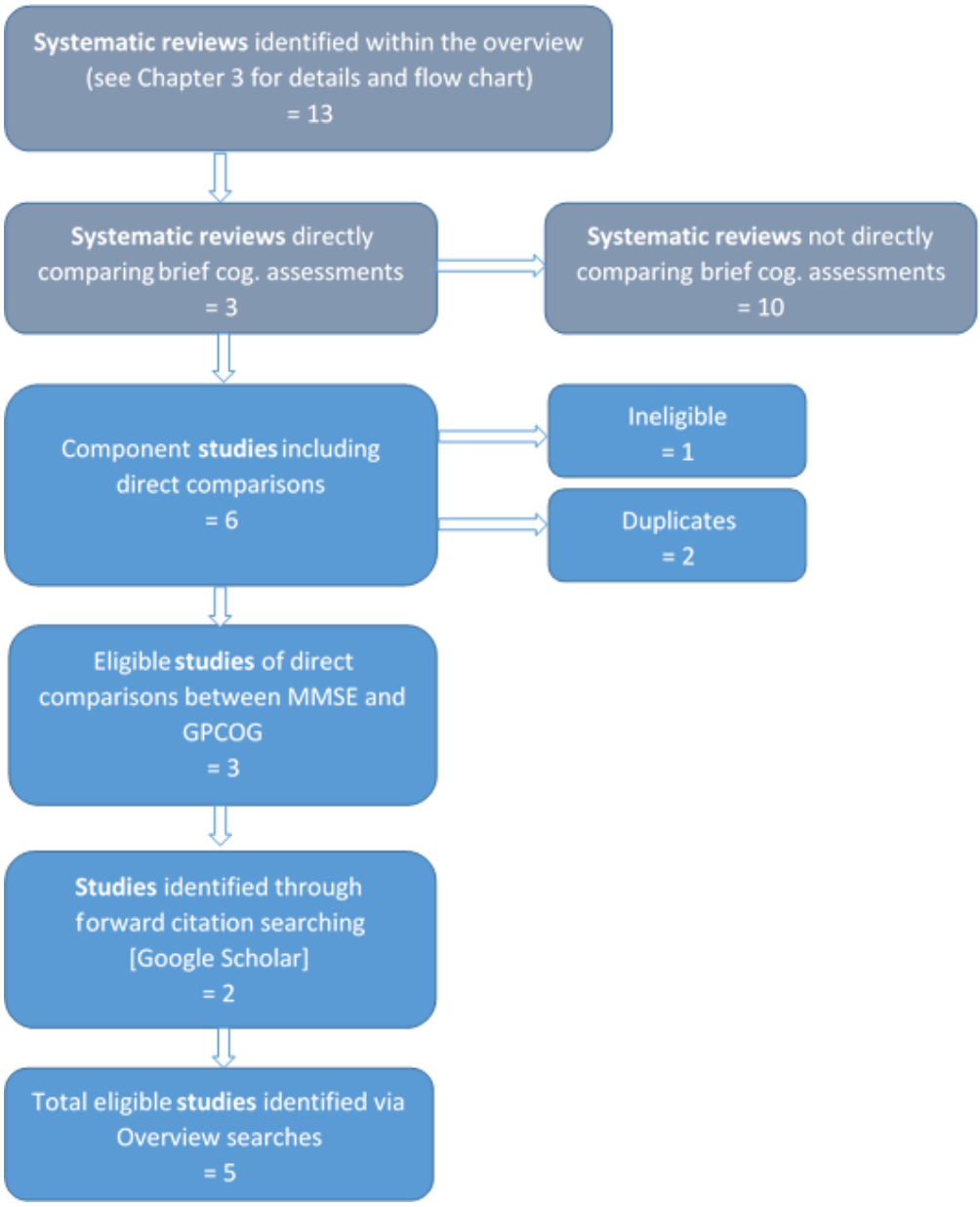
### 5.2.4. Identification of eligible studies included within the overview

The 13 systematic reviews included within the overview review[88,105,106,120,125,141,148,220-222,231,236,285] were reanalysed for all component studies that directly compared diagnostic accuracy of GPCOG and MMSE aginst one another within a primary care setting. The process of identifying eligible studies (directly comparing MMSE and GPCOG) from the 13 systematic reviews included within the overview is shown in Figure 4.

Five studies within the 13 systematic reviews included direct comparisons of more than one brief cognitive assessment. Of these, three studies [142,286,287] compared GPCOG and MMSE and were screened for inclusion in our systematic review.

Citation searching was conducted using Google Scholar to identify all indexed hits that cited these three studies. This identified two further eligible studies[43,288], giving a total of five studies identified via the overview review.

Figure 4 Flow chart of studies identified via the overview




Grey box = systematic review data; blue box = study data

### 5.2.5. Identification of novel eligible studies

The entire process for identifying novel eligible studies for this systematic review is shown in Figure 5. One further study[119] was identified via Zetoc alerts using the terms "MMSE", "GPCOG", "test accuracy" and "dementia". This study had been published since the overview was conducted, so had not been identified within the original included reviews and component studies.

In line with the published protocol[284], a standard search was run using Embase (1974 to current), PsychINFO (1806 to current) and Ovid MEDLINE (1946 to current) databases using the search terms (((Dementia and GPCOG and MMSE) or Mini

Mental State Exam) and Accuracy).af.. In order to identify new studies published since searches were conducted for the overview (see Chapter 4) the starting date parameter was set to 2015. This is one year previous to the published date of the latest included systematic review, Creavin 2016[105].

After deduplication, database searches identified a further 77 sources which were added to the 5 identified from the overview and the one study identified via a Zetoc alert. Screening notes were written to assist in decision making (shown in Appendix 9). Two screening reviewers (HH and SvK) independently piloted the first 15 hits organised alphabetically to make sure the process was understood and the notes helpful and clear. All 83 titles and abstracts were then independently screened by the same two reviewers using the online web application Rayyan QCRI[289], with 100% agreement reached on all decisions.

Seventy-six sources were excluded at title and abstract screening, with 7 studies identified for full text screening (see Table 15).

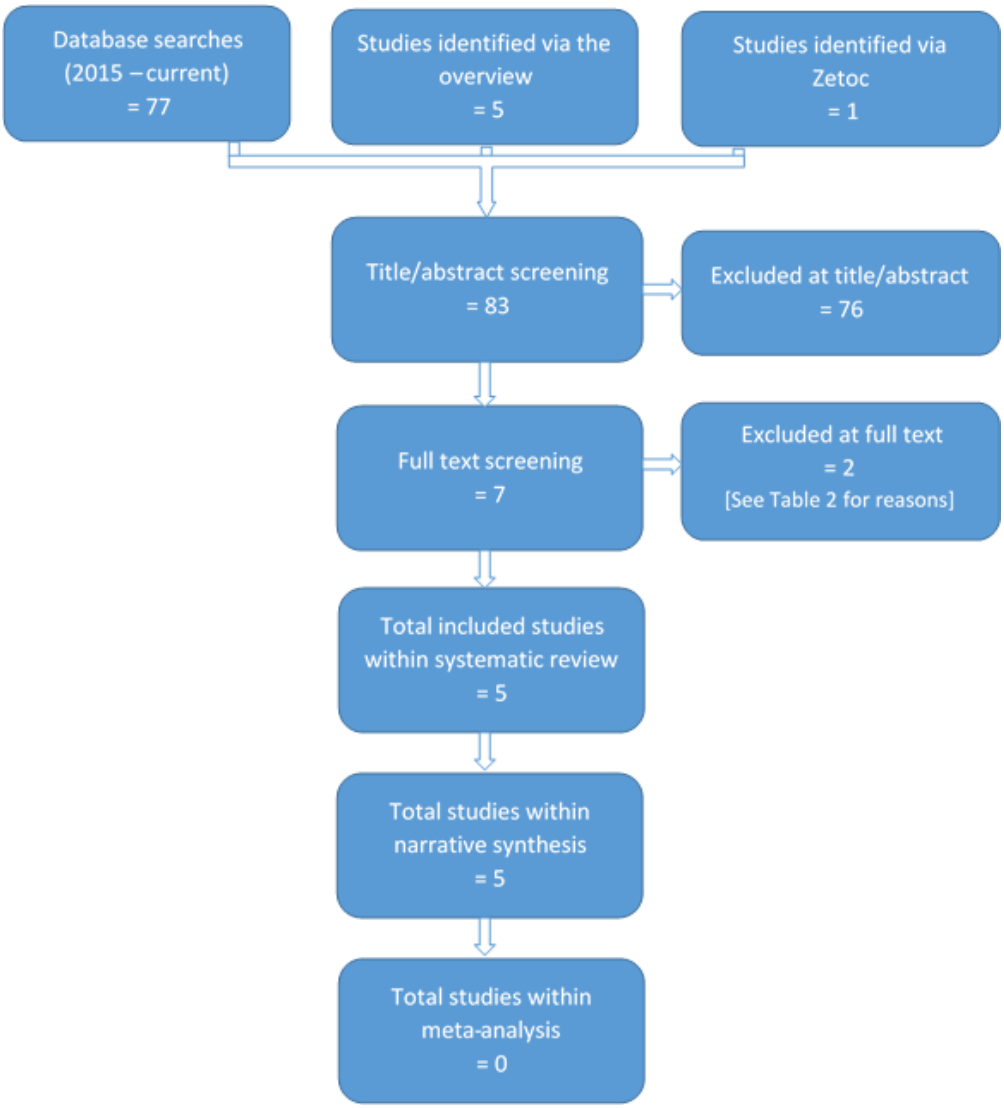
Table 15 Sources screened at full text and inclusion/exclusion decisions

| Studies screened at full text | Inclusion/exclusion decision | Reason for exclusion (if applicable) |
|---|---|---|
| **Basic 2009[287]** | Inclusion | - |
| **Brodaty 2002[142]** | Inclusion | - |
| **Brodaty 2006[236]** | Exclusion | Assessment across different populations |
| **Brodaty 2016[119]** | Inclusion | - |
| **Culverwell 2008[43]** | Exclusion | Assessment across different populations |
| **Li 2013[288]** | Inclusion | - |
| **Pirani 2010[286]** | Inclusion | - |

Two studies[43,236] were excluded at the full text stage because they assessed brief cognitive assessments within different populations, so lacked direct comparisons.

Five studies[119,142,286,288,290] were included directly comparing the diagnostic accuracy performance of MMSE and GPCOG as part of the process for identifying dementia within a primary care/general practice population.

Figure 5 Flow chart of all identified studies




### 5.2.6. Index tests

The index tests of interest were the MMSE[115] and the GPCOG[142]. The MMSE is one of the most widely used brief cognitive assessments used the within research literature[291], and development of the GPCOG has been independent to the development of the MMSE[119].

The conventional threshold for the MMSE is 24 (also shown as <24), where out of a maximum possible 30 points, scores below 24 indicate impairment. The GPCOG comprises of two sections: the section completed by the individual being assessed, known as GPCOG Patient, and an optional section for a relative or friend to complete (if present) known as GPCOG Informant. GPCOG Patient has 9 items with possible total scores of between 0 (indicating severe impairment) and 9 (indicating no impairment). GPCOG Informant has 6 items with possible total scores of between 0 (indicating severe impairment) and 6 (indicating no impairment). GPCOG Patient can

be conducted by itself, with a conventional threshold of 8 out of 9 (<8). If informants are available, a score of GPCOG Patient between 5 and 8 precipitates the GPCOG Informant and the scores are combined ("GPCOG Total") with a conventional threshold of 11 out of a maximum 15 (<11). If no informant is available, the GPCOG Informant is not completed and the conventional threshold of 8 stands. It is also possible to conduct a staged GPCOG assessment where GPCOG Informant is only required if GPCOG Patient is scored between 5 and 8 out of 9. This is known as "GPCOG Two stage".

Originally the plan was to stratify GPCOG into three types of test (GPCOG Patient, GPCOG Total and GPCOG Two stage). This was in order to maximise clinical relevance of the review, as GPCOG Informant is not used independently for assessment in the clinical setting. Once included studies had been identified, it became clear that GPCOG Informant had been assessed as an individual tool in four out of the five included studies. The GPCOG subsets were therefore included as 4 categories of test in line with the data presented within identified studies: GPCOG-informant with a threshold of >5 and <8; GPCOG-Patient with a threshold of <8, GPCOG Total with a threshold of <11 and GPCOG Two stage.

Terminology around the GPCOG subtypes also differed between studies. Two studies[286,287] referred to GPCOG/GPCOG-It Participant as 'GPCOG/GPCOG-It Cognitive'. One study[287] referred to GPCOG Total as 'GPCOG Combined'. Throughout this systematic review the conventional labels have been used, but the different terms used by authors are reflected in Table 16 for completeness.

### 5.2.7. Reference standard

There is currently no 100% accurate ("gold standard") test for identifying dementia in primary care or general practice. Because of this, the identification of an appropriate benchmark assessment ("reference standard") for assessing dementia is not a straightforward one. In addition, some reference standards which assess cognitive impairment (e.g. CAMDEX, CAMCOG) incorporate elements of the MMSE within the tool. As an example, CAMCOG is the cognitive section of the longer Cambridge Examination for Mental Disorders of the Elderly (CAMDEX) and incorporates 19 questions from the MMSE[292,293].

Reference standards were judged to be acceptable for this systematic review based upon their common identification within research literature (as identified within the overview reported in Chapter 3), and advice from clinical neuropsychologists and old age psychiatrists within the research team and Project Advisory Group. Acceptable reference standards included the following tools alone, clinical diagnosis alone or clinical diagnosis combined with one or a combination of the following assessment tools:

- Diagnostic and Statistical Manual (DSM) III/III-R/IV/IV-R,
- Clinical Dementia Rating (CDR),
- International Classification of Diseases (ICD) 10,
- Geriatric Mental State – Automated Geriatric Examination for Computer Assisted Taxonomy (GMS-AGECAT),
- Cambridge Mental Disorders of the Elderly Examination (CAMDEX),
- International Psychogeriatric Association World Health Organisation (IPA-WHO) criteria.

Generally, reference standards are selected on the basis of many variables such as common practice within individual clinics, practitioner preference, specialisation and experience of healthcare professionals and practice managers. Reference standards are also subject to changes in cost and fashion. For example, since 2001 the holders of the MMSE copyright have been enforcing a cost per use of the test as well as pursuing tests based upon the MMSE (such as the Sweet-16) though the courts, leading to debates around the use and ethics of copyright for ubiquitous and previously-free materials[265,294-296].

Many of the globally-accepted reference standards such as the World Health Organisation-supported ICD and the DSM produced by the American Psychiatric Association are updated regularly; the DSM-5 (sometimes referred to as DSM-V) was released in 2013[297], and the ICD-11 was released in 2018[298]. It was therefore decided that the criterion was valid within this wider context.

### 5.2.8. Data extraction, selection and coding

All sources were managed using EndNote X7.7 software. Two reviewers (HH and SvK) independently piloted the process for screening titles and abstracts on the first

15 sources and screening notes were produced to guide our decisions on title and abstract and full text screening. Title and abstract and full text screening were conducted independently by the same two reviewers, and a third reviewer (Chris Hyde) was available to resolve any disagreements. In the event, this was not necessary and both review authors reached consensus without contention.

A bespoke data extraction form (see Appendix 10) was piloted independently with one included study[287] by the two reviewers, and small improvements were made to the form following discussion between the reviewers. Key data extracted included characteristics of included studies (such as authors, year of study, date of publication, country of study, outcomes reported, test timings) and general limitations as well as components of the 2x2 table (TP, FP, TN, FN) or other accuracy data such as sensitivity, specificity and disease prevalence if raw numbers were not available. The data extraction form was accompanied by briefing notes. Data were abstracted by one reviewer (HH), with all extractions spot-checked by a second (SvK), and a third reviewer (CH) acting as moderator as necessary. Again, this was not needed in the circumstances as small queries were resolved by discussion between the two reviewers.

### 5.2.9. Assessment of methodological quality

The QUADAS-2[299] tool was used to assess methodological quality of diagnostic accuracy studies for systematic reviews. While this tool is developed for studies focussing on a single index test, its' suitability was assessed for studies that focus on direct comparisons of two index tests by piloting the QUADAS-2 tool on one of the included studies. QUADAS-2 uses prompts (or 'signalling questions') to explore 4 main areas of potential bias and applicability: patient selection, index test, reference standard, and flow of patients through the study and timing of the index test(s) and reference standard. The tool was tailored to fit this particular review, with a domain assessing each index test (MMSE and GPCOG) separately. Questions within Domain 2 were tailored accordingly; where 'Index tests' states: "If a threshold was used, was this pre-specified?" this was changed to simply ask "Was/were the threshold(s) pre-specified?" as both of the index tests (GPCOG and MMSE) have thresholds that should be reported.

Domain 3: Reference standard asks "Were the reference standard results interpreted without knowledge of the results of the index test?" This was amended to ask "Were

the reference standard results interpreted without knowledge of the results of both index tests?" to include consideration of both MMSE and GPCOG. Two reviewers (HH and SvK) piloted the tailored QUADAS-2 and accompanying notes on one study[287], and one reviewer (HH) quality assessed the included studies with spot checks provided by a second reviewer (SvK). An example of the QUADAS-2 form is shown in Appendix 11, and our tailored quality appraisal notes are shown in Appendix 12.

### 5.2.10. Data synthesis and analysis

Study specific estimates of the sensitivity and specificity (and their 95% confidence intervals) of GPCOG and MMSE are presented graphically via forest plots. Forest plots and summary receiver operating characteristic (SROC) plots are used to visually explore heterogeneity. Results are summarised narratively and displayed within a summary of results table (Table 20 Summary of results from included studies). Possible subgroup analyses were considered for investigating tests using lower and higher thresholds and variations in cases and control groups (e.g. confirmed dementia, probably dementia, people with memory problems, healthy people). Reporting bias was not assessed, in line with the recommendations of the Cochrane Diagnostic Test Accuracy Handbook[300]. This is because the impact of reporting bias on studies of diagnostic accuracy is unclear, and tools for its investigation are in the early stages of development[301].

## 5.3. Results

### 5.3.1. Characteristics of included studies

Key study characteristics (lead author, date of publication, country of origin, study design, setting, dementia severity, number of patients mean age plus standard deviation, participants, reference standard, index tests, thresholds and total score, other outcomes measured) across all five included studies are shown in Table 16.

*Study design*

Of the five studies[119,142,286-288] within this review, three studies employed cross-sectional designs, where all patients are recruited and receive the reference standard and index tests prior to identification of the target condition[302]. One study reported comparisons of diagnostic accuracy between MMSE, GPCOG and the Rowland Universal Dementia Scale (RUDAS) using a cross-sectional study design within a symptomatic population of 151 community-dwelling older people with early dementia[287]. One study[142] adopted a cross-sectional design to examine diagnostic

accuracy comparisons between GPCOG, MMSE and the Abbreviated Mental Test (AMT) within a mixed symptomatic and asymptomatic population of 283 community dwelling patients. One study[119] used a cross-sectional study design to compared the diagnostic accuracy of GPCOG and MMSE in a mixed symptomatic and asymptomatic population of 2,028 community dwelling patients, although the precise mix of symptomatic versus asymptomatic participants and how they were identified and allocated is unclear in the reporting of this study.

Two studies used two-gate case-control study designs[286,288], where groups of participants with and without the target condition are identified before the index test is performed[302]. One study[288] explored diagnostic accuracy of the Chinese version of GPCOG (GPCOG-C), MMSE and Hasegawa's Dementia Scale (HDS) in a sample comprising 253 symptomatic community-dwelling volunteers with concerns about their memory (controls) and 103 psychogeriatric clinic outpatients with subjective memory complaints (cases). One study[286] examined the diagnostic accuracy of an Italian version of GPCOG (GPCOG-It) compared to the MMSE using a case-control design in a sample of 182 community-dwelling patients aged 55 years and over with subjective memory complaints (cases) and 78 patients without memory complaints identified for inclusion by their GPs (controls). For the purposes of this systematic review, results of the RUDAS, AMT and HDS were excluded.

*Study populations*

Recruitment was from the community and general practice, with one study[288] also recruiting from a hospital and the psychogeriatric department of a local mental health centre. Li 2013[288] and Brodaty 2002 [142] excluded participants who had signs of delirium, as well as those with visual or hearing impairments. Basic 2009[287] also excluded on this basis but added an exclusion for people with physical impairments. Brodaty 2002 also excluded people with a diagnosis of depression or if they had poor English language abilities. Brodaty 2016[119] excluded people with neurological disease, psychotic symptoms, developmental disabilities, substance abuse, progresive malignancy, or an illness that the GP judged may impede study completion[119]. Pirani 2010 excluded all patients with known cognitive impairment, or comorbidity "predisposing to cognitive disorders such as metabolic and cardiovascular disorders" (p.83[286]).

*Recruitment*

Recruitment of study participants was from the community and general practice, with one study[288] also recruiting from a hospital and the psychogeriatric department of a local mental health centre.

*Setting, timings and blinding*

Participants were recruited from various locations (see above), but all were assessed by aged care psychiatrists, geriatricians or senior registrars in specialist geriatric or psychiatric disciplines[286-288] except in two studies where follow up assessments were carried out by research psychologists[142] or research nurses[119].

In one study there was a lack of clear separation in process between administration of the reference standard and index tests. Within the case control study by Pirani 2010, the patient group ('cases') were recruited over 6 months either after referral to a Dementia Assessment Unit (DAU) by GPs "trained in administering the GPCOG-It in their daily practice" (p.83[286]), or self-referred to the DAU seeking medical advice. The study flow diagram makes it clear that GP referral was based upon results of the GPCOG-It. The control group were recruited by GPs from their "well known attendees without known cognitive impairment" (p.84) and referred on to the DAU without undergoing the GPCOG-It. At one month follow up, GPCOG-It was administered to those who had not undergone previous assessment (i.e. self-referred cases or GP-recruited without symptoms) alongside MMSE, CAMCOG and ADAS-Cog assessment. Dementia diagnosis was made by expert geriatricians blinded to GPCOG-It scores using a semi-structured interview, physical exam and results of MMSE, CAMCOG and ADAS-Cog assessments. The MMSE results used to compare against the GPCOG-It were derived from this process, meaning the MMSE assessment formed both part of the reference standard and the index test. As referred to in the Methods section (5.2.6), Cambridge Cognitive Examination (CAMCOG)[119] also incorporates the MMSE. The Alzheimer's Disease Assessment Scale – Cognitive section (ADAS-Cog) was designed to measure aspects of cognition in Alzheimer's disease including components on memory language and praxis[303] and appears to have been developed independent of both MMSE and GPCOG.

Whilst the assessment of GPCOG-It in this study appeared to be independent of the reference standard, the assessment of MMSE clearly was not independent. Therefore

the ability of the MMSE to accurately identify people with cognitive impairment may have been overestimated or underestimated due to incorporation bias.

Li 2013 used GPs or junior psychogeriatricians to assess participants using the GPCOG-C, MMSE and HDS. Following this (with timings between assessments unreported), one of four senior geriatricians blinded to the earlier results interviewed participants and their informants in order to make a diagnosis of dementia using DSM-IV and clinical judgment. Part of this process involved assessment of the participant's cognitive and functional abilities, but no detail was provided within the paper on how this was done or whether a cognitive assessment tool was used.

Within the 2002 study by Brodaty[142], following recruitment GPs administered an unrefined version of GPCOG plus the AMT to consecutive patients and contacted an informant known to the patient for at least 5 years, either by telephone or in person. Presumably this was in order to conduct the Informant section of the GPCOG, although this is not stated within the paper. Around 5 weeks later, a research psychologist visited the participant at home and carried out assessments including the CAMDEX and GPCOG with an informant interview where possible. It was not reported whether the research psychologist was blinded to previous assessment results. The CAMDEX incorporates the MMSE (as part of the cognitive subscale CAMCOG) and so, as with Pirani 2010, this introduces a potential source of incorporation bias where the CAMDEX is used as part of the reference standard as was the case in this study.

Basic 2009[287] recruited participants from memory clinics to which they had been referred by GPs, themselves and relatives, or community aged care teams. Before recruitment, they were assessed for cognition by a senior geriatrician, aged care psychiatrist or senior registrar with geriatric or psychiatry specialty and were then assessed by a research assistant at their home, the clinic another location. Timings between assessments are not reported within the manuscript, and no details of blinding are provided. One other brief cognitive assessment (the RUDAS) was conducted by a research assistant blinded to scores, so the inference could be that other assessments were not blinded as otherwise this would have been noted. Data from both the MMSE and GPCOG were used to decide DSM-IV diagnoses, again introducing the potential for incorporation bias which may result in an overestimation of effect.

*Study size*

Basic 2009[287] reported the smallest study sample size with 151 participants, in a study using a cross-sectional design. Participants were 'older' (age limits not given), community-dwelling with mixed symptomatology. Pirani 2010[286] included a sample size of 200 participants using a case-control design consisting of 132 people within a symptomatic patient group and 68 people in an asymptomatic control group. Brodaty 2002[142] employed a cross-sectional study design which included 283 participants of mixed symptomatic and asymptomatic presentation, with asymptomatic patients being aged 75 years and over, and symptomatic patients suspected of having a memory problem aged between 50 and 74 years old. Within this study there were discrepancies amongst sample sizes presented, with different sample sizes reported across all measures (GPCOG Patient: N=282, GPCOG Two Stage: N=246, MMSE: N=283 and AMT: N=269) except GPCOG Informant and GPCOG Total: N=202). The only explanation offered by the authors was that sample sizes varied due to missing data (see Table 1 in the original study[142]). Li 2013[288] included 356 participants within a double-gate case-control design consisting of 103 people who were outpatients of a psychogeriatric clinic with memory complaints (cases) and 253 people living in the community aged over 50 years old with memory concerns (symptomatic controls).

Finally, Brodaty 2016[119] included 2028 participants in a Single-gate cross-sectional study design. Participants were living in the community with mixed symptomatic and asymptomatic presentation, although the exact mix of symptoms and allocation within the study is unclear from the manuscript. Within this study, all participants did not complete all the brief cognitive assessment measures; 2028 participants were administered the MMSE, but only 1717 undertook GPCOG with missing GPCOG data of 311 participants unaccounted for. For this reason, these data have been analysed separately within this thesis as this discrepancy is not explained within the study text.

*Study authors and locations*

One study author (Henry Brodaty, developer of GPCOG) was lead author on two of the included studies[119,142] and co-author on another two of the included studies[286,288]. Three of the studies[119,142,287] were based in three states (Victoria, New South Wales, South Australia) across Australia, with one study[288] based in Shanghai, China and one study[286] based in Modena and Cento, Italy.

*Index tests*

MMSE, GPCOG, GPCOG-C, GPCOG-It, RUDAS, AMT, and HDS were all employed as index tests across included studies. As the focus of this systematic review is solely MMSE and GPCOG (or national adaptations), results of RUDAS, AMT and HDS are reported (see Table 16) but were not included in the analysis.

GPCOG subtypes were given slightly different names across studies. Basic 2009[287] refers to GPCOG Cognitive. Pirani 2010[286] similarly refers to GPCOG-It Cognitive. GPCOG and GPCOG-It Cognitive in these studies correspond to GPCOG-C Patient referred to in Li 2013[288], and GPCOG Participant referred to in the most recent included study, Brodaty 2016[119].

GPCOG/GPCOG-It/GPCOG-C consistently refer to GPCOG Informant for the informant section of the tool. GPCOG Total, however, which combines the scores of GPCOG Participant and GPCOG Informant, is referred to by Basic 2009[287] as GPCOG Combined whereas all other studies refer to GPCOG Total.

GPCOG Two-stage is only referred to in these terms across all four studies in which it features.

*Diagnostic thresholds*

MMSE was assessed at three different diagnostic thresholds (<24, <25 and <27) across different studies: one study based in Italy[286]used both <25 and <27 as thresholds for dementia. The study authors reported <25 as the 'standard' threshold, and <27 as approved under Italian law to identify people eligible for free cholinesterase inhibitors under a programme titled "Progetto Chronos".

The greatest number of included studies comparing MMSE at a single threshold and GPCOG was three[119,287,288], which reported an MMSE threshold of <24 compared to GPCOG Total and this is arguably the most clinically-relevant comparison. These distributions are illustrated in Table 17.

*Reference standards*

Reference standards differed between studies, with four studies using DSM-IV criteria[119,286-288] and one using CAMCOG[142]. One study used DSM-IV without reference to other tools[288]. Three studies used DSM-IV and at least parts of the MMSE[119,286,287]. Two of these studies used DSM-IV, MMSE and CAMCOG (which incorporates the MMSE and is part of the wider CAMDEX[293]).

As detailed in 'Setting, timings and blinding' above, CAMCOG incorporates the MMSE as part of the measure. DSM-IV decisions in one study explicitly used MMSE and GPCOG scores as part of the decision-making process[287] and two studies used MMSE as part of the DSM-IV assessment[119,286]. Only one study solely referred to using the DSM-IV[288] as a reference standard. ICD-10 criteria did not feature in any of the five included studies.

Table 16. Summary of characteristics of included studies

| Source (year) *Country* [sd; pop] | Setting | Dementia severity (measure) | # patients, $\bar{x}$ age yrs ± SD | Participants | Reference standard | Index test, threshold indicating impairment/ total score | Other outcomes measured |
|---|---|---|---|---|---|---|---|
| **Basic (2009) *Australia* [SG; symp]** | CALD community | Early dementia (Lawton IADL & GDS) | 151, 77.1 ± 8.9 | Community dwelling older people (no age limit given) | DSM-IV using MMSE and GPCOG | MMSE, <24/30<br>GPCOG Cognitive†, <8/9<br>GPCOG Informant, <5/6<br>GPCOG Combined‡, <11/15<br>RUDAS, <23/30 | MBI; GDS; Lawton IADL |
| **Brodaty (2002) *Australia* [SG; unclear]** | PC doctors' offices | NR | 283, 79.6 ± 6.1 | Community dwelling people 50-74yrs with MC; and patients ≥75 yrs | DSM-IV using CAMDEX, CAMCOG and MMSE | MMSE, <24/30<br>GPCOG Participant, <8/9<br>GPCOG Informant, <5/6<br>GPCOG Total, <11/15<br>GPCOG Two-stage, >5 <8/15<br>AMT, <8/10 | GDS; SFHS |
| **Brodaty (2016) *Australia* [SG; mixed]** | GP practices | NR | 2028, 81.1 ± 4.12 | Community dwelling people ≥75 yrs who had visited their GP within the last 24 months | CAMCOG | MMSE, <24/30<br>GPCOG Total, <12/15<br>GPCOG Two-stage, >5<8/15 | GDS |
| **Li (2013) *China* [2G; 'healthy']** | PC facility, MH centre and community | NR | 356, 72.5 ± 8.9 | People aged 50-90 yrs from the community, attending Waitan Hospital or the psychogeriatric department of the Shanghai MH Center | DSM-IV | MMSE, NR<br>GPCOG-C Patient, <8/9<br>GPCOG-C Informant, <5/6<br>GPCOG-C Total, <11/15<br>GPCOG-C Two-stage, >5 <8/15<br>HDS, <21/32.5 | Administration time |
| **Pirani (2010) *Italy* [2G; symp]** | Community & GP practices | NR | 200, 76.1 ± 7.2 | People aged over 55 yrs with an available informant referred to DAUs by GPs | DSM-IV using CAMCOG, MMSE, and ADAS-Cog | MMSE, <25/30<br>MMSE, <27/30<br>GPCOG-It Cognitive†, <8/9<br>GPCOG-It Informant, <5/6<br>GPCOG-It Total, <11/15<br>GPCOG-It Two-stage, >5 <8/15 | Administration time; dementia severity (not explicitly assessed *a priori*) |

sd = study design; pop = population; SG = single-gate cross-sectional design; 2G = two-gate case-control design; symp = symptomatic [controls]; unclear = unclear symptomatic/asymptomatic; mixed = symptomatic & asymptomatic participants; 'healthy' = healthy controls, without memory complaints; MMSE = Mini Mental State Examination; GPCOG = General Practitioner Assessment of Cognition; NR = not reported; PC = primary care; CALD = culturally & linguistically diverse; MH = mental health; yrs = years; SD = standard deviation; GPs = general practitioners; CACT = community aged care teams; MCs = memory clinics; Lawton IADL = Lawton Instrumental Activities of Daily Life scale; GDS = Geriatric Depression Scale; DSM-IV = Diagnostic and Statistical Manual version 4; MBI = Modified Barthel Index; MC = memory complaints; SFHS = short form health survey; AMT = Abbreviated Mental Test; GPCOG-C = General Practitioner Assessment of Cognition – Chinese version; HDS = Hasegawa's Dementia Scale; SMC = subjective memory complaints; DAUs = Dementia Assessment Units; GPCOG-It = General Practitioner Assessment of Cognition - Italian version; †GPCOG Cognitive = GPCOG Participant; ‡GPCOG Combined = GPCOG Total.)

Table 17. Distribution of included studies reporting on GPCOG and MMSE at different thresholds

| | MMSE | | | GPCOG | | | |
|---|---|---|---|---|---|---|---|
| | <24 | <25 | <27 | Informant | Patient | Total | Two-stage |
| **Basic 2009** | ✓ | | | ✓ | ✓ | ✓ | |
| **Brodaty 2002** | | ✓ | | ✓ | ✓ | ✓ | ✓ |
| **Brodaty 2016** | ✓ | | | | | ✓ | ✓ |
| **Li 2013*** | ✓ | | | ✓ | ✓ | ✓ | ✓ |
| **Pirani 2010†** | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

* contain comparisons between MMSE<24 and GPCOG-C, the Chinese version of the GPCOG

† contains comparisons between MMSE<25, MMSE<27 and GPCOG-It, the Italian version of GPCOG

The distribution of included studies examining GPCOG and MMSE at different thresholds is shown in

GPCOG-Total was measured in all five of the included studies, and GPCOG Informant and Patient were measured in the same four studies[142,286-288]. GPCOG Two-stage was measured in four studies[119,142,286,288].

Table 17. It is most informative to read across the table to see which studies report at the same MMSE threshold *and* the same GPCOG subsection. For example, Basic 2009[287], Brodaty 2016[119] and Li 2013[288] all reported MMSE <24, but of these all three only reported GPCOG Total scores. Only two[287,288] out of these studies reported GPCOG Informant, Patient and Two-stage sections (one of which[288] actually reported GPCOG-C, the Chinese variant). At the conventional MMSE threshold of <25, two studies[142,286] measured GPCOG Informant, Patient, Total and two-stage (although one of those studies[286] used the Italian variant of GPCOG, GPCOG-It). Only one study[286] directly compared MMSE at two different thresholds (<25 and <27) against GPCOG Total Italian version (GPCOG-It Total).

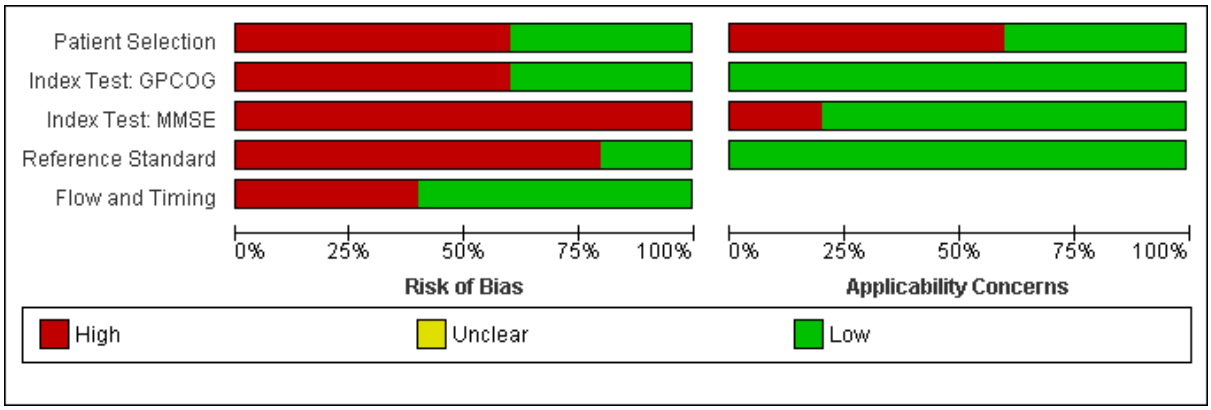
GPCOG-Total was measured in all five of the included studies, and GPCOG Informant and Patient were measured in the same four studies[142,286-288]. GPCOG Two-stage was measured in four studies[119,142,286,288].

### 5.3.2. Quality assessment

Study quality was assessed using a tailored version of QUADAS-2[299], following coproduction of quality assessment notes by Harriet Hunt and Sanne Van Kampen. These notes are shown in the appendix (Appendix 12). The QUADAS-2 tool tailored for this review and accompanying notes were independently piloted on Basic 2009[287]

by two reviewers (HH and SvK) and we compared our assessments and experiences of the process after this. The quality of the remaining four studies[119,142,286,288] was assessed, and another reviewer (SvK) spot-checked assessments. From these assessments we produced both a summary graph (Figure 6) and a table of judgements across domains for each included study (Figure 7) to illustrate these quality assessment decisions.

Figure 6. QUADAS-2 Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies



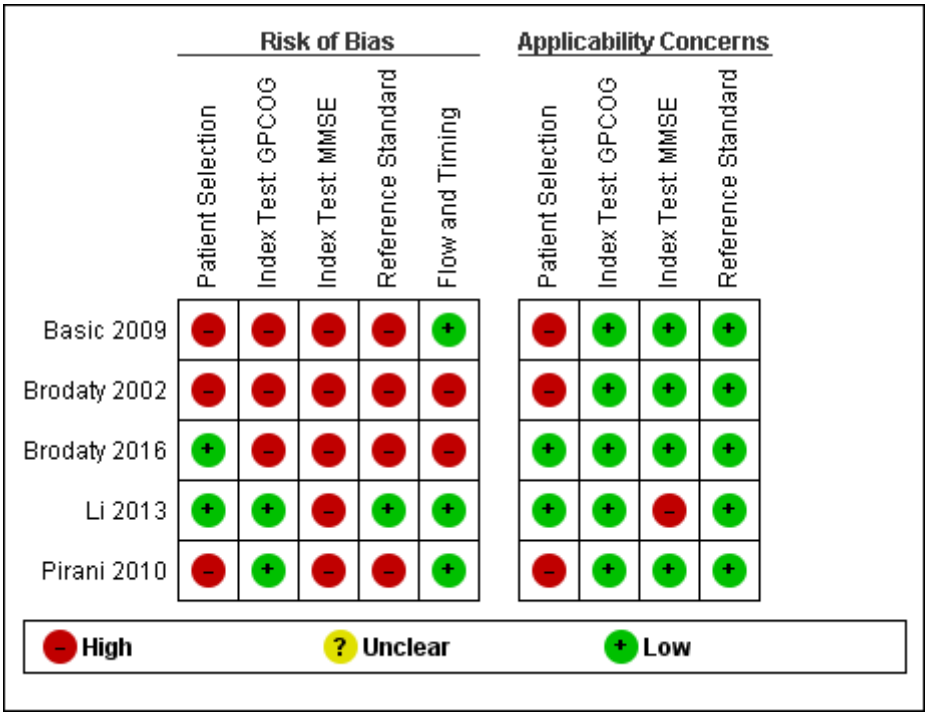

As shown in Figure 6, we assessed patient selection as introducing a moderate to high (50-75%) risk of bias when summarised across all five studies. Two studies[286,288] employed two-gate case-control designs, with Li 2013 including symptomatic controls with memory concerns, and Pirani 2010 including “healthy” controls recruited from routine general practice excluding symptomatic patients and those with conditions “predisposing to cognitive disorders”.

A challenge of the QUADAS-2 instrument is that, whilst case-control designs are highlighted as a potential source of bias, there is no further advice or recommendation on how to then handle these studies. There is also no assessment of the type of case-control design studies use, for example the sampling strategies employed and whether they are single-gate (from the same sample) or two-gate (from two different samples). These approaches vary considerably and introduce different types of biases and potential variables depending on the target population, recruitment, and sampling framework. If a single sample is used, one might reasonably expect less variation than if two distinct samples are recruited as in a two-gate study design. Test accuracy will vary across different subsets of participants, so particular attention should be given to diagnostic accuracy studies with two different sampling schemes for symptomatic cases and healthy controls, as was the case with the case control designs included in

this review. One major issue of systematic reviews including studies of direct comparisons is the potential for spectrum effect, where highly-selected symptomatic participants, or conversely an overly-narrow range of 'healthy' participants, can overestimate any observed differences in diagnostic accuracy. Equally, the presence of alternative conditions in 'healthy' individuals, and the presence of comorbidities in either group, can introduce systematic bias as a result of the study design used. These factors are not explored in the current version of QUADAS-2, but another iteration of this tool for assessing the quality of diagnostic accuracy studies would benefit from specifically considering these scenarios.

Risk of bias concerns for use of GPCOG as an index test were similarly moderate to high when summarised across studies. Three of the included studies (Basic 2009[287], Li 2013[288] and Pirani 2010[286]) blinded raters of GPCOG to results of the reference standard. Two of the included studies (Brodaty 2002[142] and 2016[119]) did not blind raters of the GPCOG to results of the reference standard or the other index test (MMSE, which in four cases [119,142,286,287] formed both an index test *and* reference standard, either in isolation or as part of the CAMCOG).

Figure 7. QUADAS-2 Risk of bias and applicability concerns: review authors' judgements about each domain for each included study



When summarised across all studies (see Figure 6), risk of bias concerns for use of the MMSE as an index test were high (100%). This was primarily due to interpretation of the index test taking place without blinding to the results of the reference standard. In four studies, the reference standard incorporated the MMSE index test either as part of the CAMCOG[119,286], the broader (and encompassing) CAMDEX[142] or simply as part of the composite tools used by assessors to arrive at a DSM-IV diagnosis[287].

Whilst not specifically assessed within QUADAS-2, Li 2013[288] did not prespecify the MMSE threshold used within the study, meaning optimum diagnostic thresholds could have been selected *post-hoc*. Introducing the potential for data-driven selection of optimal thresholds is a significant concern[304], because such discrimination can lead to overly-optimistic assessments of sensitivity and specificity. This risk is heightened in studies with smaller sample sizes as is the case with Li 2013[288] (N=253).

Risk of bias concerns around the reference standard were also high (75-100%) when summarised across all five included studies. This was due to the potential for incorporation bias in using an index test (MMSE) as part of the reference standard in four studies[119,142,286,287], either as a discrete tool or as part of the CAMCOG. One study (Li 2009[288]) solely used a psychogeriatricians' clinical judgment based upon DSM-IV criteria as the reference standard, which we judged to contribute a low risk of bias to reference standard assessments.

A moderate risk of bias (25-50%) was found for flow and timing summarised across all studies. This was negatively skewed by two studies, Brodaty 2002[142] and 2016[119], where all participants were not included in the analysis and were recorded as missing data without explanation or discussion within the manuscripts. Within Brodaty 2016, 311 participants were recorded as "missing GPCOG data" meaning the total samples of participants differed between those assessed using MMSE (n=2028) and those assessed using GPCOG (n=1717). These missing data were explicitly reflected within the study flow diagram but not elsewhere. This constitutes a loss of 15% of participants (2028/1717), distributed asymmetrically and solely within the GPCOG sample.

Within Brodaty 2002, 81 patients were missing between recruitment and administration of the MMSE to conduct of the GPCOG, although with no study flow diagram provided it is unclear at what stage the participants went missing. This constitutes a loss of 29% of participants (283/202) and again, is distributed

asymmetrically within the GPCOG sample. These are not small losses, and this substantial percentage of missing data alongside the asymmetry of the loss indicated a 'high' risk of bias rating for flow and timing. These losses were not initially obvious within the manuscript and only emerged during synthesis, and so these ratings were revised from an initial rating of 'low' risk of bias. This raises additional issues for quality assessment ratings generally and handling of the data, which will be addressed in more detail within the Discussion section of this chapter.

In terms of applicability of the individual studies to the research question "How accurate are GPCOG and MMSE in identifying dementia within a general practice setting when directly compared to each other?", patient selection was assessed as being particularly problematic with a moderate to high (50-75%) concern of applicability. Basic 2009[287] and Pirani 2010[286] excluded patients with delirium or visual, hearing or physical impairments[287] and known cognitive impairment or comorbidity related to cognitive disorders including metabolic and cardiovascular diseases[286]. Similarly, Brodaty 2016[119] excluded patients with neurological disease, psychotic symptoms, developmental disability, substance abuse, progressive malignancy, or an illness that the general practitioners judged may prevent patients completing the study. Brodaty 2002[142] excluded patients with delirium, depression, poor English language abilities, sight or hearing. Li 2013[288] excluded participants with acute or unstable psychiatric disorders "such as major depression", as well as delirium, anxiety, or a vision or hearing impairment. This study also excluded people without a suitable informant available. It was concluded that these exclusions meant the patient group assessed were less suited for a typical general practice population, where multimorbidities and concurrent impairments would not be unusual[51].

Applicability of the MMSE as index test was assessed as a low to moderate concern (25-50%). Four studies were rated as a low concern except Li 2013[288] which was rated as a high concern. In this study, the authors did not pre-state the MMSE threshold used, and so raised doubts over whether post-hoc selection of an optimum threshold had been introduced.

There were no notable concerns in applicability to the research question for the remaining domains, so in these remaining areas all studies were rated as low concerns.

Brodaty 2002 was judged as a high risk of bias across all five domains (Patient selection, index test: GPCOG, index test: MMSE, reference standard, and flow and timing). These ratings were due to the non-randomised recruitment of participants and the number of potentially-inappropriate exclusions such as the exclusion of participants with clinically suspected dementia but without corroborative history to assess decline. Index tests and reference standards were measured without blinding of the assessors, and one of the index tests (MMSE) was incorporated into the reference standard as part of the CAMCOG, which in itself was part of the CAMDEX tool used to derive criteria for a DSM-IV-based diagnosis of dementia. There was no evidence that the MMSE threshold was prespecificed, and there was reference to a post-hoc analysis of the MMSE at a lower threshold of <24. GPCOG threshold was pre-specified, but was administered twice to all participants with a 5-week wait in between assessments. This repeat testing was in order to refine the measure, but did mean participants underwent the same test within a relatively short period of time which may have introduced further confounding. As referenced earlier, missing data was a major issue in this study with 29% of participants missing between recruitment and conduct of the MMSE to conduct of GPCOG.

In terms of individual study quality, Brodaty 2016 was rated as a low risk of bias in one domain (Patient selection), with a high risk of bias rating across four domains - the index test: MMSE, index test: GPCOG, reference standard, and flow and timing of the study. Applicability concerns were low across all domains, meaning we assessed the study results as broadly applicable to our study question. This was a particularly problematic assessment. The low risk of bias in the 'Patient selection' domain was difficult for handling missing data, as the QUADAS-2 question relates to patient selection and recruitment, rather than patient flow but this distinction feels artificial compared to practical data handling. The flow and timing question does not ask about patients receiving the index tests, neither is there any question nor prompt to determine whether, if all patients were not included in the analysis, the missing data were reasonably accounted for. In terms of applicability concerns, recruitment seemed reasonable given the limited data available within the manuscript, and broadly matched the population in question.

The QUADAS-2 question asks *'Are there concerns that the index test, its conduct, or interpretation differ from the review question?"* to which the answer in this case was

no, because the review question was clearly stated in terms of the population, index tests, condition and outcomes so searches were designed specifically to address this question. Answering 'yes' to this question would imply that searches were poorly-designed or conducted, and/or that the review protocol was inappropriate for answering the research question. Both Basic 2009[287] and Pirani 2010[286] were rated as high risk of bias and high applicability concerns in patient selection as they used double-gate case-control designs, which have been shown to exaggerate overall diagnostic accuracy[305].

*Translated versions*

Two of the included studies included variants of GPCOG adapted for languages other than English; one variant is translated into in Italian, the GPCOG-It[286] and one into Chinese, the GPCOG-C[288]. Both measures were formed as a translation of the original GPCOG[142] maintaining the original format and scoring system. Both adapted measures were developed by research teams including the originator of GPCOG, and both measures maintained all aspects of the test except the source language.

*Rater differences*

There were differences between the raters of the included measures. None of the included studies incorporated test assessors who were explicitly identified as general practitioners, although all studies stated the test of interest would be for use in a general practice or community population. In the study by Li 2013[288], six out of ten of their raters were psychogeriatricians who were "well trained and experienced in dementia screening" with between eight and 20 years' experience of screening practice. In this study, raters used clinical judgment rather than any pre-specified diagnostic criteria.

In the study by Basic 2009[287], raters were aged care clinicians and not explicitly GPs – although these assessors were potentially closest to the target rater population. In the study by Brodaty 2002[142], the tests were administered by research psychologists and in Brodaty 2016 [119] both GPCOG and MMSE were administered by trained research nurses. Finally, the study by Pirani 2009 [286] included test administration by expert geriatricians or neurologists rather than GPs.

*Incorporation bias*

There was significant potential for incorporation bias across four of the five included studies, where elements of one or more of the index tests under assessment are

included within the reference standard, thus risking overestimation of agreement between measures. In Basic 2009, data from both MMSE and GPCOG were used to arrive at DSM-IV assessments. Brodaty 2002 used a modified version of the CAMDEX as their reference standard, which contains elements of the MMSE. Brodaty 2016 used CAMCOG as the reference standard, which again contains elements of MMSE. Pirani 2010 included a reference standard of DSM-IV assessment based upon a semi-structured interview with both patient and informant, a physical examination, as well as scores on Italian versions of MMSE, CAMCOG and ADAS-COG. Only Li 2013 used a reference standard of clinical judgment that was assessed by raters (geriatricians) blinded to MMSE, GPCOG-C and HDS scores.

### 5.3.3. Comparing sensitivity and specificity

The Forest plot in Figure 8 shows sensitivity, specificity and confidence intervals across all included tests (GPCOG Informant, Patient, Total and Two Stage) as well as MMSE at three thresholds (<24, <25 and <27).

In GPCOG Total, the one measure where all five included studies[119,142,286-288] report data, sensitivity ranges from 0.79 to 0.98, with specificity ranging from 0.78 to 0.92. The balance between sensitivity and specificity scores also vary without a consistent direction of effect, with Basic 2009 reporting higher sensitivity of 0.98 (95% CI 0.91-1.00) and lower specificity of 0.78 (95% CI 0.68-0.86) and Brodaty 2016 reporting lower sensitivity of 0.79 (95% CI 0.71-0.86) and higher specificity of 0.92 (95% CI 0.91-0.93).

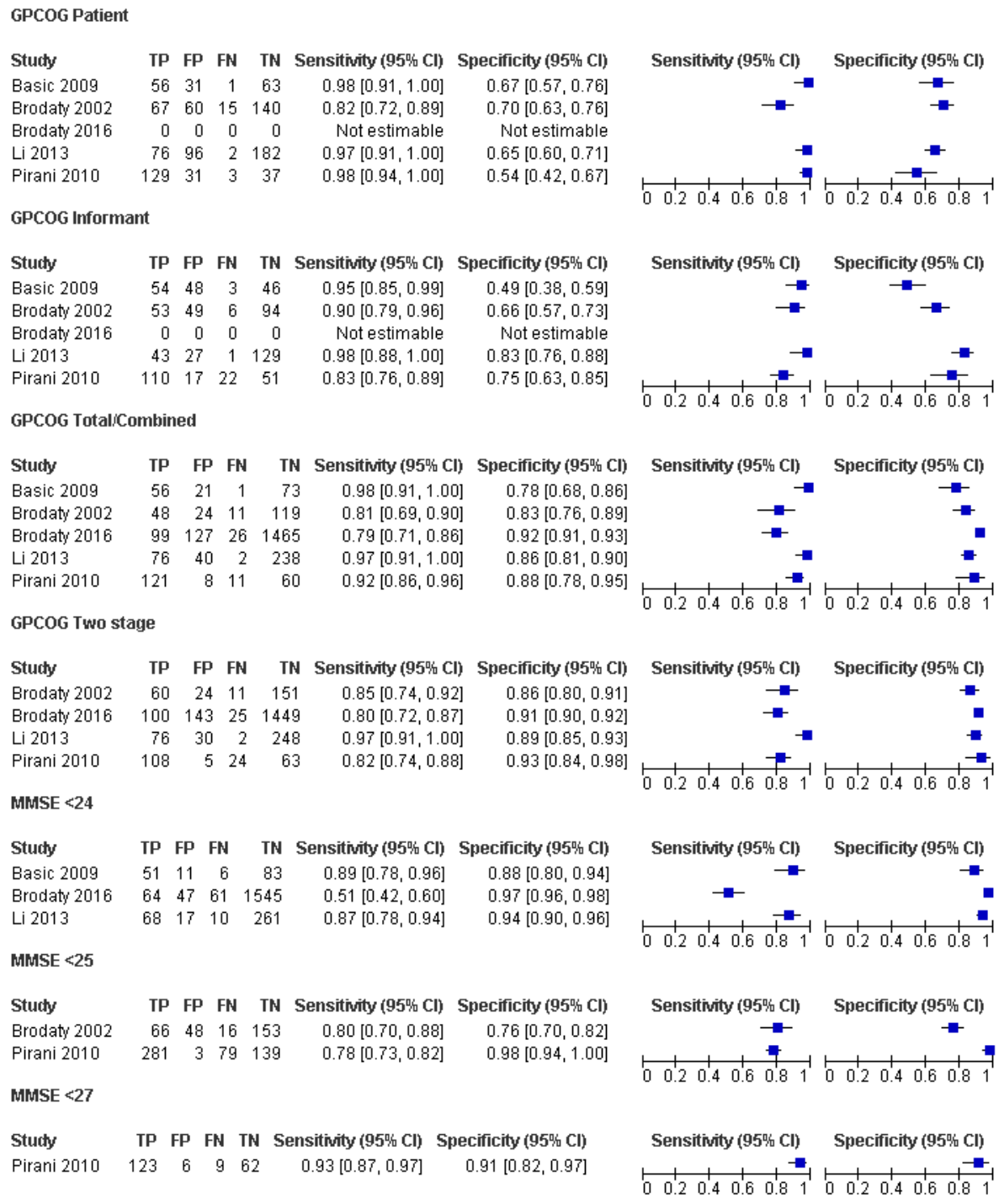
GPCOG Informant accuracy measures show consistent direction of effect, with the four smaller studies[142,286-288] which included this measure all reporting higher sensitivity (0.83-0.98) and lower specificity (0.49-0.83). GPCOG Patient performance shows the same pattern across the same four studies, with higher sensitivity (0.82-0.98) and lower specificity (0.54-0.70).

GPCOG Two-stage accuracy reveals similar variability to GPCOG Total, with the four studies[119,142,286,288] reporting a mix of higher sensitivity (0.97, 95% CI 0.91-1.00) and lower specificity (0.89, 95% CI 0.85-0.93) in Li 2013 compared to lower sensitivity (0.80-0.85) and higher specificity (0.86-0.93) in the other three studies.

MMSE accuracy measures across thresholds demonstrated similar variability, with MMSE <24 measured in three studies ranging from sensitivity 0.51-0.89 and specificity

0.88-0.97. MMSE <25 was without clear direction across two studies[142,286] reporting sensitivity 0.78-0.80 and specificity 0.76-0.98. At the highest threshold of MMSE <27, Pirani 2010 reported sensitivity of 0.93 (95% CI 0.87-0.97) and specificity of 0.91 (95% CI of 0.82-0.97).

Figure 8. Forest plot displaying sensitivity and specificity of all brief cognitive assessments across studies




*N.B.* 0.1 sensitivity difference from that reported in Brodaty 2002[142] due to rounding up in the original

paper – recalculated here using actual figures. Brodaty 2002 and 2016 measures of MMSE and GPCOG report imbalanced sample sizes.

GPCOG Total was assessed in all five included studies[119,142,286-288], MMSE <24 was assessed in three of the included studies[119,287,288] and MMSE <25 was assessed in two [142,286] out of the five included studies.

These measures are clinically relevant for comparison in terms of being used individually within a general practice setting as part of the process for assessing cognitive function, so may be fairly compared to one another (as opposed to, for example, MMSE versus the GPCOG Informant which is unlikely to be used in isolation within clinical practice). MMSE thresholds of <24 and <25 are both commonly used in practice and research, with their relative strengths and weaknesses for a general practice population not yet certain [65,105,106,119,146,265,306]. Therefore it would seem to be of benefit to assess the directly comparable accuracy of these two MMSE thresholds separately, in order to identify any variation in sensitivity or specificity that would suggest one threshold to have benefits over the other in specific situations (e.g. if ruling out dementia was the clinical priority).

### 5.3.4. GPCOG Total versus MMSE<24

Table 18 illustrates that at an MMSE threshold of <24, there was an absolute difference between GPCOG and MMSE sensitivity of 8.77% in Basic 2009, and 10.26% in Li 2013. The absolute difference in specificity between MMSE <24 and GPCOG Total was -10.64% in Basic 2009, and 1.57% in Li 2013.

At an MMSE threshold of <25, the absolute difference in sensitivity was 0.87% for Brodaty 2002 and 13.64% for Pirani 2010. The absolute difference in specificity for these measures was 7.10% in Brodaty 2002 and 10.25% in Pirani 2010, as shown in Table 18.

Table 18. Sensitivity and specificity comparing GPCOG Total versus MMSE thresholds <24 and <25

| Study | Sensitivity [true positives/total cases] | | Difference (95% Confidence Interval) | Specificity [true negatives/total non-cases] | | Difference (95% Confidence Interval) |
|---|---|---|---|---|---|---|
| | GPCOG Total | MMSE<24 | | GPCOG Total | MMSE<24 | |
| **Basic 2009[287]** | 0.98 [56/57] | 0.89 [51/57] | 8.77 (-0.62 to 19.47) | 0.78 [73/94] | 0.88 [83/94] | -10.64 (-21.31 to 0.19) |
| **Brodaty 2016[119]** | 0.79 [99/125] | 0.51 [76/148] | 27.85 (16.64 to 37.95) | 0.92 [1465/1592] | 0.97 [1824/1880] | -5.00 (-6.59 to -3.49) |
| **Li 2013[288]** | 0.97 [76/78] | 0.87 [68/78] | 10.26 (1.75 to 19.64) | 0.86 [238/278] | 0.94 [261/278] | -8.27 (-13.41 to -3.25) |
| | GPCOG Total | MMSE<25 | | GPCOG Total | MMSE<25 | |
| **Brodaty 2002[142]** | 0.81* [48/59] | 0.80* [66/82] | 0.87 (-12.86 to 13.50) | 0.83 [119/143] | 0.76 [153/201] | 7.10 (-1.71 to 15.33) |
| **Pirani 2010[286]** | 0.92 [121/132] | 0.78 [103/132] | 13.64 (5.03 to 22.23) | 0.88 [60/68] | 0.98 [67/68] | -10.29 (-20.14 to -1.74) |

* Reported in Brodaty 2002[142] as sensitivity 0.82 and 0.81, due to rounding up in the original paper. N varies due to missing data - Brodaty 2002 and 2016 report imbalanced sample sizes.

Within the studies included in this systematic review, GPCOG Total demonstrated consistently higher sensitivity compared to MMSE<24, at the expense of reductions in specificity.

This is similarly illustrated in the Summary Receiver Operating Characteristic (SROC) plot shown in Figure 9. This SROC shows paired accuracy of GPCOG Total and MMSE summarised across the studies that reported these data, in contrast to the study level ROC plots presented and discussed further in this chapter which simply present individual study-level sensitivity and specificity scores.

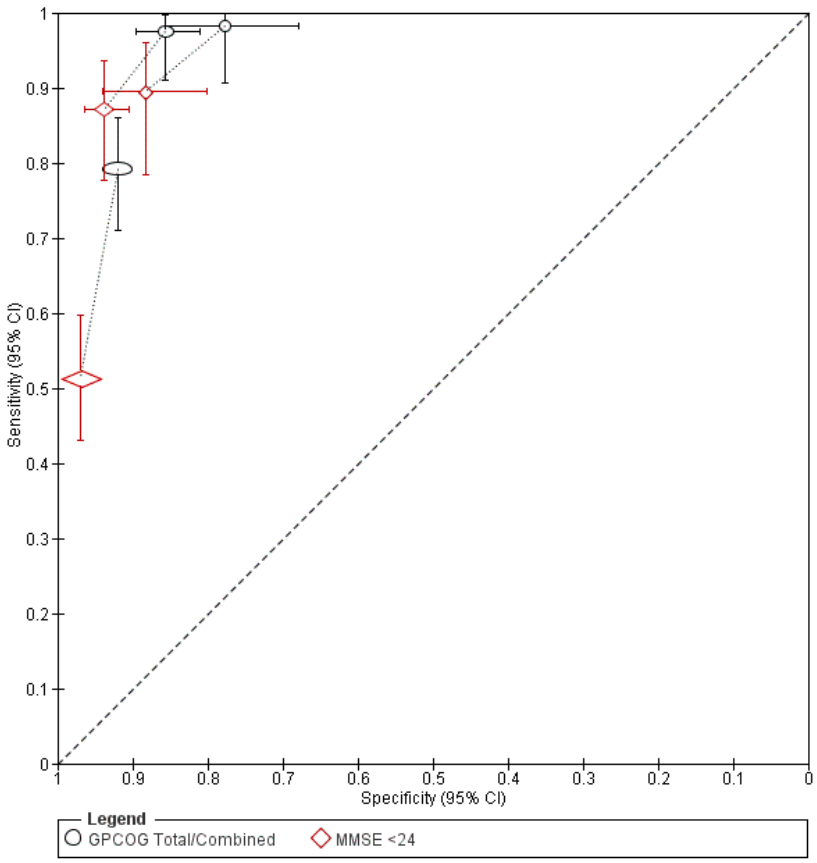
As shown in Table 18, accuracy data reported within two studies[287,288] directly comparing MMSE<24 and GPCOG Total shows a mixed relationship between these two tests.

As referenced earlier, both Brodaty 2002 and Brodaty 2016 reported different sample sizes for each index test. Brodaty 2002 reported a sample size of 202 for GPCOG Total, and 283 for MMSE <25 with no clear reason given for the loss of 29% of the recruited sample. Brodaty 2016 reported 1717 for GPCOG and 2028 for MMSE, and whilst the samples seem to include the same participants, as reported within the study

flow diagram, there is no explanation for the loss of 311 people between tests within the original study paper. Whilst this population is relatively large which accounts for the lower confidence intervals, this is still an unaccounted loss of 15% of the recruited study sample, distributed asymmetrically between tests (the loss was in the GPCOG arm) and it is difficult to say what effect this may have on results. This study reported lower sensitivity across both MMSE <24 and GPCOG Total, and specificity was higher in both tests. The gap between sensitivity and specificity was most marked in MMSE <24 where sensitivity was near chance at 0.51 [95% CI, 0.43-0.60] and specificity was 0.97 [95% CI, 0.96-0.98]. With GPCOG Total, sensitivity was 0.81 [95% CI, 0.69-0.90], whereas specificity was 0.92 [95% CI, 0.91-0.93].

When direct comparisons of GPCOG and MMSE are viewed within the same paired accuracy SROC plot the comparative performance of individual tests is plainly illustrated, as shown in Figure 9.

Figure 9 Paired accuracy of GPCOG Total and MMSE SROC plot from studies that reported MMSE at the <24 threshold [Basic 2009, Brodaty 2016 and Li 2013]



The SROC plot in Figure 9 illustrates the paired sensitivity and specificity of GPCOG Total compared to MMSE from studies that reported MMSE accuracy at the <24 threshold (Basic 2009[287], Brodaty 2016[119] and Li 2013[288]) with MMSE shown as a red diamond and GPCOG Total as a black circle.

The size of the diamond or circle represents the number of participants within the study, with Brodaty 2016 having the largest sample (N=2028/1717 – note the unaccounted discrepancy) and Basic 2009 having the smallest sample (N=151). There is an observable pattern of higher sensitivity in the GPCOG Total and higher specificity of MMSE across the three studies, suggesting that there may be a trade-off between the higher sensitivity of GPCOG Total and marginally higher specificity of MMSE – although this is based upon small samples within imperfect studies, so results should be treated with caution.

### 5.3.5. GPCOG Total versus MMSE <25

Only two of the included studies (Brodaty 2002[142] and Pirani 2010[286]) reported MMSE diagnostic accuracy at a threshold of <25 compared to that of GPCOG Total. Within the study by Brodaty 2002[142], at an MMSE threshold of <25, GPCOG Total demonstrated superiority in both sensitivity and specificity as shown in Table 18. Higher sensitivity scores were statistically significant in Pirani 2010[286], as judged by a lack of overlap in 95% confidence intervals.

Paired comparisons between GPCOG Total and MMSE reported by Brodaty 2002[142] and Pirani 2010[286] illustrated in Figure 10 reveal a similar pattern in the set of paired accuracy scores from Pirani 2010 to those observed in data reported at the MMSE<24 threshold shown in Figure 9 – but again, a small number of samples and variable study quality means any inferences drawn are highly limited.

One set of paired accuracy scores from Brodaty 2002, however, appears to be an outlier. As illustrated in both Table 18 and Figure 10, GPCOG Total and MMSE accuracy scores appear to be similar, with reported GPCOG Total sensitivity of 0.81 and specificity of 0.83, and MMSE sensitivity of 0.80 and specificity of 0.76 (N.B. these sensitivity figures differ by 0.1 from the sensitivity scores reported in the original paper, due to rounding up by the study authors). This lower specificity of 0.76 reported in the MMSE accuracy score is at odds with the other included studies, where MMSE specificity ranged from 0.88 to 0.98. The GPCOG Total specificity of 0.83 is the middle

range of GPCOG Total specificity scores reported within the other included studies, where GPCOG Total specificity ranged from 0.78 to 0.92.

In Brodaty 2002[142], the authors suggest two main reasons for MMSE and GPCOG accuracy displaying similar performance. Firstly, rater bias may have been present as test administrators were clinical psychologists rather than general practitioners. Secondly, MMSE formed part of the DSM-IV-based reference standard so there was circularity of process which may have inflated the MMSE accuracy performance. Whilst these two sources of potential bias are present in this study, they are also present in other included studies. Within Li 2013[288], tests were administrated and rated by a combination of GPs and psychogeriatricians, Brodaty 2016[119] used trained research nurses, Pirani 2010[286] used geriatricians, neurologists and psychologists to administer the tests, and Basic 2009[286] used research assistants and aged care clinicians for test administration. As described under 'Quality assessment', in four of the included studies[119,142,286,287] the reference standard incorporated the MMSE index test either as part of the CAMCOG[119,286], CAMDEX[142] or as part of the composite tools used by assessors to arrive at a DSM-IV diagnosis[287].

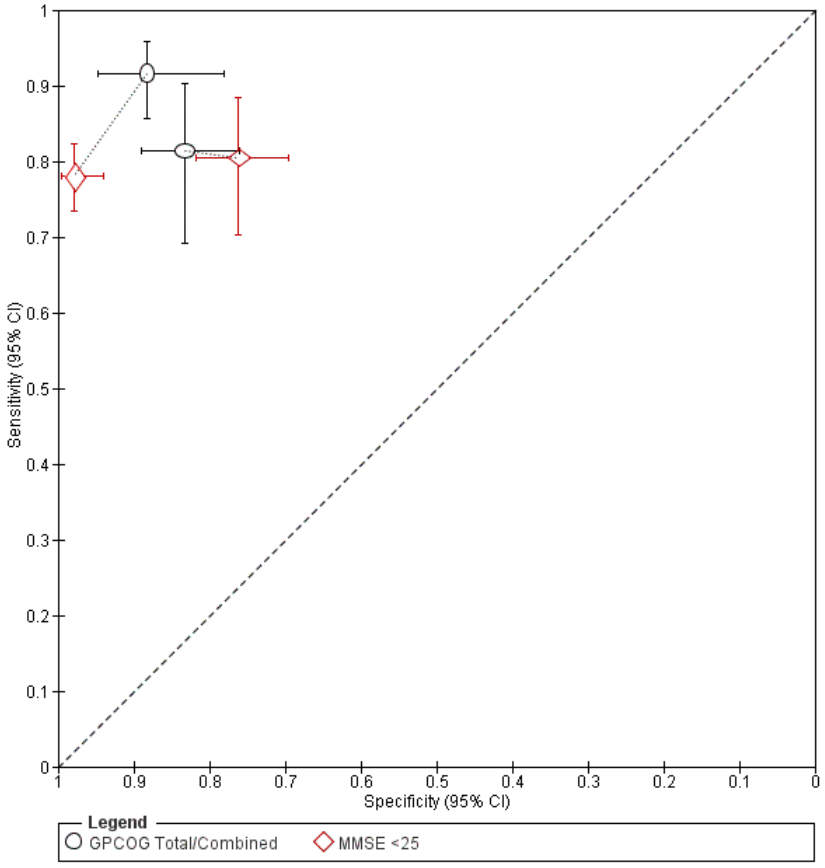
As these sources of potential bias are present in both Brodaty 2002 and Pirani 2010, it may be fairly hypothesised that any biases would be operating similarly on test performance across each of the tests. Therefore observable differences are less likely due to these potential biases, as they would be equally present across all studies vulnerable to these sources of bias and so hold a degree of symmetry across included studies.

This may not be the case with incorporation bias, where the potential bias introduced may affect different aspects of test performance differently. It could therefore be suggested that incorporation bias is asymmetric, as in a topic area such as where the reference standard is a composite measure formed of e.g. neuropsychological assessment and clinical judgment based on international disease classifications, the degree to which an index test is "incorporated" into a reference standard is not a binary result and will vary across studies.

In the present systematic review (where in four out of the five included studies the MMSE has been used as *at least part of* the reference standard), this may well mean that scores on the MMSE as index test will carry greater congruence with the reference

standard (against which it is being assessed) and higher sensitivity and specificity will result. The degree to which incorporation bias is present and has the potential to influence results of the study varies between included studies, however carefully the study parameters have been decided and prespecified.

Figure 10 Paired accuracy of GPCOG Total and MMSE SROC plot from studies that reported MMSE at the <25 threshold [Brodaty 2002 and Pirani 2010]




This SROC plot illustrates inconsistent accuracy scores of MMSE across studies, with GPCOG Total displaying slightly higher sensitivity in both studies and MMSE showing varied specificity. This would suggest two things: primarily, two studies are insufficient to show a clear direction of accuracy between these two brief cognitive assessments; secondly, that the specificity of MMSE may vary depending on the influence of other factors.

### 5.3.6. Comparing Receiver Operating Characteristic (ROC) curves

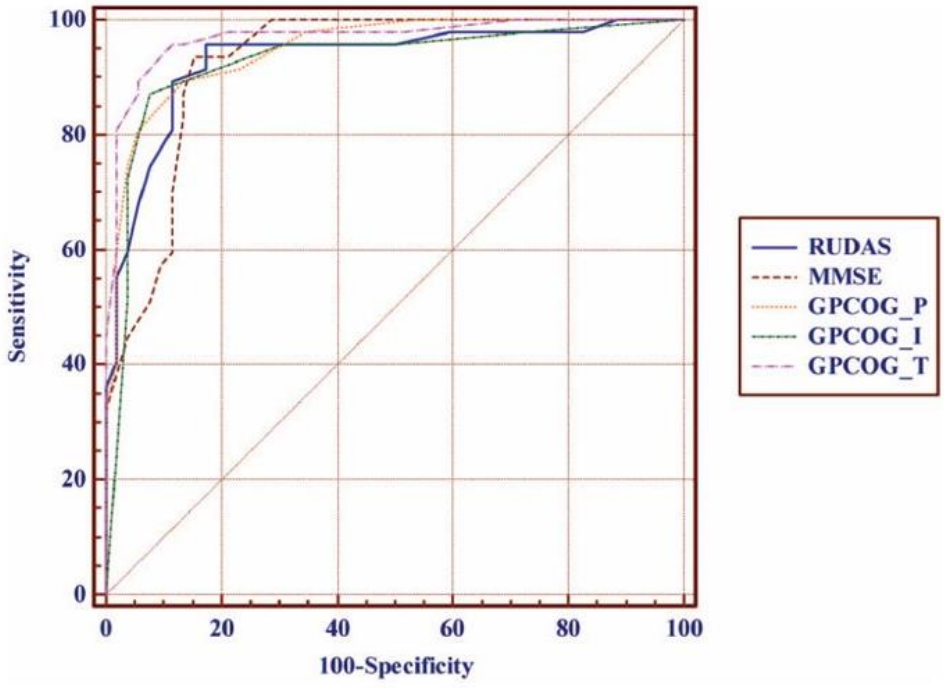
For every individual test threshold, there is a pair of sensitivity and specificity values which can be plotted on a graph. Conventionally this graph displays sensitivity-1 on the x axis, and specificity on the y axis - although as can be seen in figures 8-12 below

this convention is not always followed. The further into the top left hand corner of the graph a curve is drawn, the higher the sensitivity and specificity of the test being plotted. Optimum thresholds can be viewed within the context of interactions between sensitivity and specificity, although these thresholds may not always be practicable depending on the test in question. When assessing optimal thresholds for a clinical setting, we should interpret ROC curves within this wider context[307].

The five individual ROC curves within included studies are shown below in figures 8 to 12, reproduced with the kind permission of the study authors. The individual study ROC plots are described, and then discussed in summary presenting Area Under the Curve (AUC) data in a table and re-plotted as a summary ROC (SROC) curve. It is worth noting that, currently, the only way to include ROC curves in a systematic review is to reproduce the figures used to generate the curves and then re-plot them.

Basic 2009 created an ROC curve plotting three different tests (RUDAS, MMSE and GPCOG) with the GPCOG reporting three different measures shown in Figure 11. The focus has not been on RUDAS (the solid line in the ROC curve) as this was not a measure of interest for this systematic review.

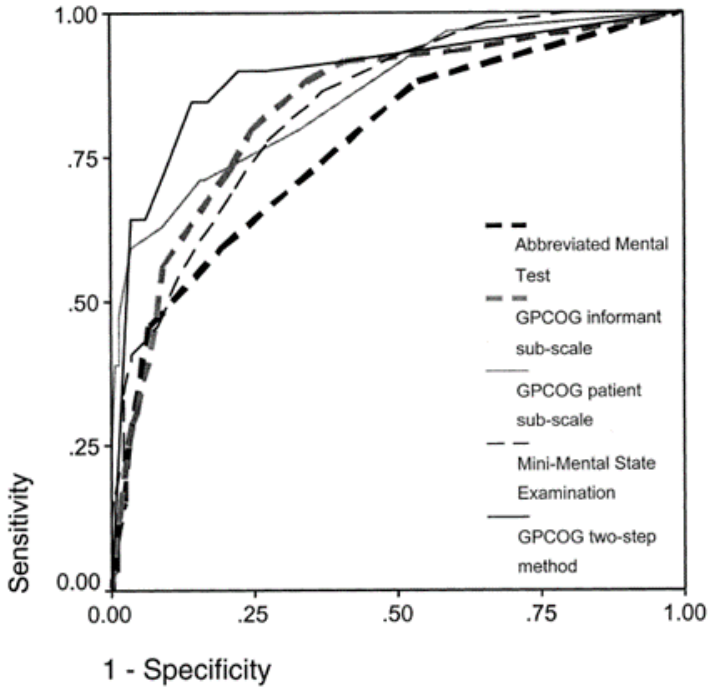
Figure 11. Basic 2009 ROC curve comparing RUDAS, MMSE, GPCOG Patient, Informant & Total




Reproduced with kind permission of the study authors.

According to the ROC curve in Figure 11, MMSE displays similar sensitivity and specificity to GPCOG Patient and GPCOG Total until a point on the curve when MMSE sensitivity and specificity dip lower than GPCOG. MMSE, GPCOG Patient and GPCOG Informant converge at the mid-point of the curve, whilst GPCOG Total maintains a higher curve on both sensitivity and specificity until MMSE reaches and maintains higher sensitivity with falling specificity. None of the measures display a clear and consistently superior ROC curve within this plot.

Figure 12 Brodaty 2002 ROC curve comparing AMT, MMSE, & GPCOG Patient, Informant and Total




Reproduced with kind permission of the authors.

Brodaty 2002 created an ROC curve plotting three different tests (AMT, MMSE and GPCOG) with the GPCOG reporting four different measures. GPCOG Two-Step is reflected in the ROC curve, but it uses a binary outcome of patients scoring above 5 but below 8 out of a possible 9. In their later paper[119] the authors highlight their decision *not* to represent GPCOG Two Step on the ROC curve in Figure 15 “because it provides a dichotomous outcome rather than a continuous range of scores” [119] (p.327).

In According to the ROC curve in Figure 11, MMSE displays similar sensitivity and specificity to GPCOG Patient and GPCOG Total until a point on the curve when MMSE sensitivity and specificity dip lower than GPCOG. MMSE, GPCOG Patient and

GPCOG Informant converge at the mid-point of the curve, whilst GPCOG Total maintains a higher curve on both sensitivity and specificity until MMSE reaches and maintains higher sensitivity with falling specificity. None of the measures display a clear and consistently superior ROC curve within this plot.

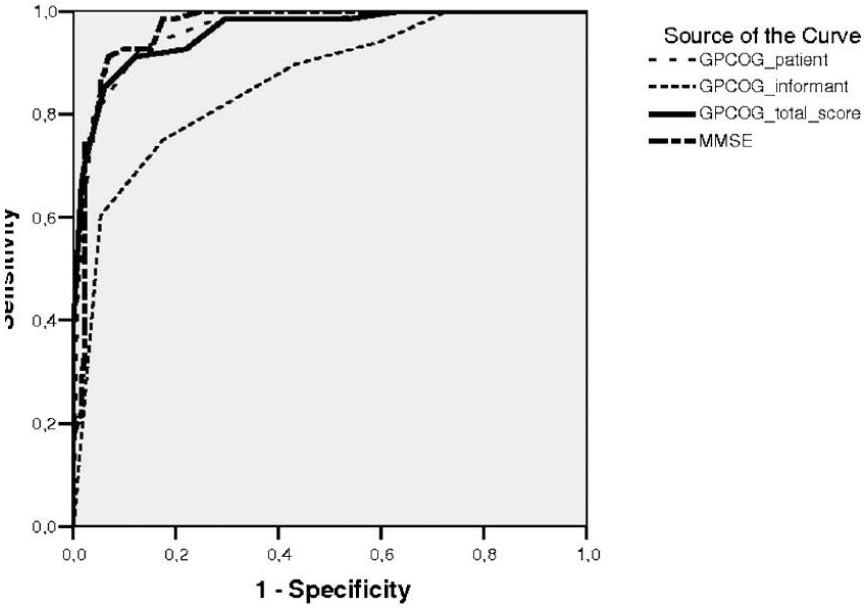
Figure 12, the lines of interest are the MMSE which is the finest dashed line, the GPCOG Informant represented by a thick dashed line, GPCOG Patient represented by a thin continuous line, and GPCOG Two Step, which is a thick continuous line. As with Basic 2009, AMT results (shown as a thick dashed line on the ROC curve) were discounted as this test was not a focus of this systematic review.

This curve shows slight variation between the two tests at different points, not dissimilar to those observed in the ROC plot produced by Basic 2009 (see According to the ROC curve in Figure 11, MMSE displays similar sensitivity and specificity to GPCOG Patient and GPCOG Total until a point on the curve when MMSE sensitivity and specificity dip lower than GPCOG. MMSE, GPCOG Patient and GPCOG Informant converge at the mid-point of the curve, whilst GPCOG Total maintains a higher curve on both sensitivity and specificity until MMSE reaches and maintains higher sensitivity with falling specificity. None of the measures display a clear and consistently superior ROC curve within this plot.

Figure 12) and there is little obvious overall superiority of one test over another.

Pirani 2010 created an ROC curve plotting MMSE and GPCOG, with the GPCOG reporting three measures. Figure 13 shows the ROC curve produced by Pirani 2010.This ROC curve shows that MMSE accuracy (thick dashed line) performs similar to GPCOG-It Total (thick solid line), with minimal variation between these measures as observed in the ROC curves for MMSE and GPCOG produced by Basic 2009, Brodaty 2002 and Brodaty 2016.

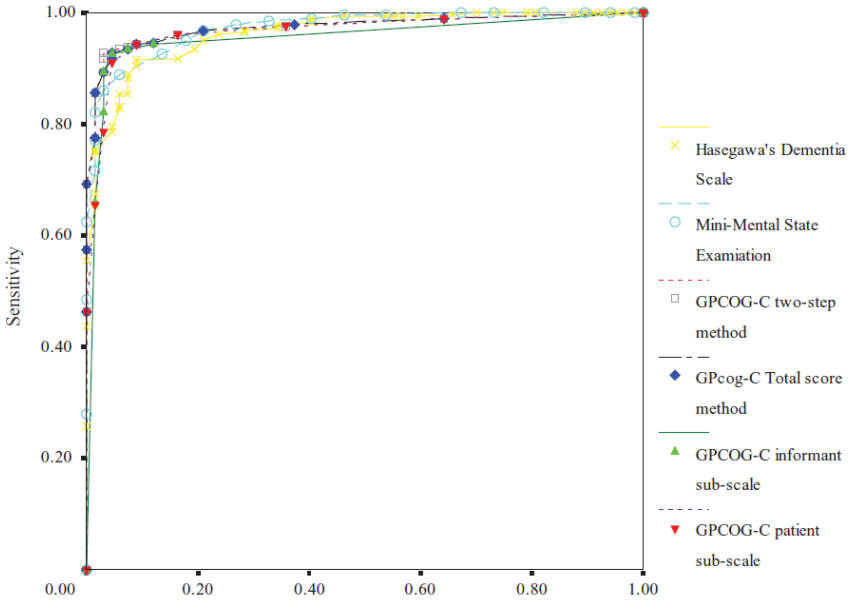
Figure 13. Pirani 2010 ROC curve comparing MMSE, GPCOG-It Patient, Informant & Total




Reproduced with kind permission of the authors. Legend from study authors states GPCOG rather than GPCOG-It.

Li 2013 created an ROC curve shown in Figure 14 comparing three different tests (HDS, MMSE and GPCOG-C) with GPCOG-C plotted across 4 different measures. GPCOG-C Two-stage (also known as GPCOG-C Two-Step) was also plotted on the ROC curve but, similar to Brodaty 2002, no threshold is reported as it uses a binary outcome so it is unclear what scores were used to plot the ROC curve. MMSE is plotted on the ROC curve created by Li 2013, but the authors do not state which threshold was used. There is no focus on HDS (the yellow line with asterisk scores marked on the ROC curve) as this was not a measure of interest for this systematic review.

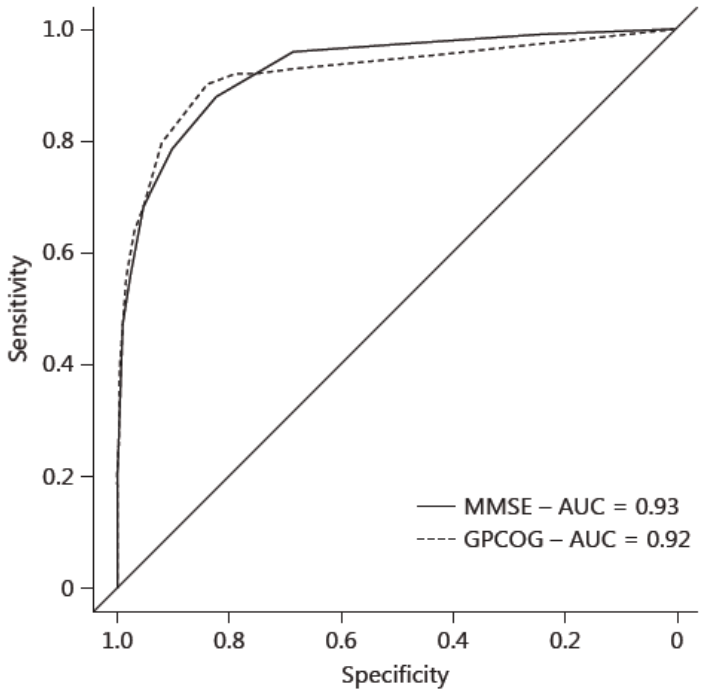
Figure 14. Li 2013 ROC curve comparing Hasegawa's Dementia scale, MMSE, and GPCOG Patient, Informant, Two Step and Total




Reproduced with kind permission of the authors

This ROC curve shows that MMSE (regular dashed line with circles) performs similar to GPCOG-C Total (irregular dashed line with diamonds), with minimal variation between these measures as observed in the ROC curves displaying MMSE and GPCOG produced by Basic 2009, Brodaty 2002, Brodaty 2016 and Pirani 2010.

Figure 15. Brodaty 2016 ROC curve comparing MMSE & GPCOG Total




Reproduced with kind permission of the authors.

Brodaty 2016 created a ROC curve shown in Figure 15, plotting MMSE and GPCOG Total. These results are displayed with the caveat that sample sizes between the two tests differed without clear explanation, and as this may have contributed an unknown degree of bias to the results. GPCOG Two-Step was not reflected in the ROC curve as the outcome is dichotomous rather than a set of continuous variables which could be plotted on an ROC curve. As see in Basic 2009 and Brodaty 2002, both curves follow similar trajectories and differences are minimal, with no clear superiority of one test over another.

These ROC plots display little variation in curve shape across the five included studies.

In theory, it is possible to compare global diagnostic performance by observing the Area Under the Curve (AUC), which can vary from 1.0 (a perfect diagnostic test) to 0.5 (a non-discriminatory diagnostic test). Whilst AUC gives no information about sensitivity and specificity, it can indicate overall accuracy of a diagnostic test[307]. Reported AUC of four included studies are shown in Table 19.

Table 19 Comparing AUC across GPCOG Total and MMSE

| | GPCOG (total) | | MMSE | | Effect |
|---|---|---|---|---|---|
| | AUC | 95% CI | AUC | 95% CI | |
| **Studies reporting MMSE main threshold as <24** | | | | | |
| **Basic 2009** | 0.97 | 0.91, 0.99 | 0.93 | 0.88, 0.97 | GPCOG better |
| **Li 2013** | 0.97 | 0.96, 0.99 | 0.97 | 0.96, 0.99 | No difference |
| **Brodaty 2016** | 0.92 | 0.89, 0.95 | 0.91 | 0.89, 0.94 | GPCOG better |
| **Studies reporting MMSE main threshold other than <24** | | | | | |
| **Brodaty 2002** | 0.91 | 0.86, 0.95 | 0.85 | 0.80, 0.90 | GPCOG better |
| **Pirani 2010** | 0.96 | 0.93, 0.98 | 0.96 | 0.93, 0.98 | No difference |

In studies that reported diagnostic accuracy of MMSE at the threshold of <24, differences in performance based on ROC curves between GPCOG Total and MMSE were marginal and not significant in two studies[119,287] with no observable difference in a third study[288]. In studies that reported diagnostic accuracy of MMSE at other thresholds, ROC curves illustrate that differences were marginal and not significant.

### 5.3.7. Other comparative data

Whilst other comparative data were recorded wherever it was found within included studies, there was little additional reporting of common factors across studies.

Ancillary accuracy data were not identifiable from published studies, and whilst aggregate disease severity data were reported, specific information was lacking on

the severity of missed diagnoses in false negatives, or whether false positives were free of all disease. It was not possible to identify whether there was concordance between false negatives and false positives identified by each test.

There was also a lack of additional information which may have proved useful to help general practitioners to make a decision on the best brief cognitive assessment to use and would be feasible to collect in a comparative accuracy study. In particular, data on administration time and acceptability of the test to patient, carers and primary care team would prove invaluable for decision makers when considering how available brief cognitive assessments compare to one another in a clinical setting.

Only two of the included studies empirically addressed other aspects of the test performance[142,286] and clearly reported data from which conclusions were derived. Brodaty 2002[142] was the only included study which reported a GP satisfaction survey, and concluded that the GPCOG Two stage was efficient in terms of time, with the majority of GPs surveyed on test satisfaction (N=67) rating the GPCOG as being practical [87.8%], economically viable [87.8%] and acceptable to patients [98%]. The study authors report test administration time for the GPCOG Total as taking “less than 4 minutes” with the informant section administered in “less than 2 minutes” (p.533). Pirani 2010 was the only included study to report actual administration time for the included tests, and is worth noting to highlight the absence of this information within the other included studies. Reported administration time (with standard deviations in minutes) required for GPCOG-It was significantly lower than for other cognitive scales (p.88): GPCOG-It Patient = average 2.97 minutes (Standard Deviation (SD)±0.20); GPCOG-It Informant = average 1.53 minutes (SD±0.22); GPCOG-It Total score = average 4.45 minutes (SD±0.42); MMSE = average 8.9 minutes (SD±1.1). It is unclear who measured and assessed administration time, and they may not have been general practitioners - leaving questions about generalisability of these results to a general practice population and applicability of findings to a clinical setting.

### 5.3.8. Subgroup analyses

Subgroup analyses were considered for investigating tests using lower and higher thresholds and variations in cases and control groups (e.g. confirmed dementia, probably dementia, people with memory problems, healthy people). In the event, there were insufficient data to conduct subgroup analyses across more than one study in many cases. This decision was taken in line with our published protocol[284].

## 5.4. Main findings

### 5.4.1. Which test is more accurate?

When directly comparing the diagnostic performance of MMSE and GPCOG (including GPCOG-It and GPCOG-C) within the same study, which test is more accurate in identifying people with possible dementia? Based on the limited evidence of mixed quality from five studies identified within this systematic review, MMSE and GPCOG Total perform similarly across both sensitivity and specificity when used to identify patients with possible dementia within a general practice setting.

### 5.4.2. Optimum threshold selection

Thresholds may have more influence on accuracy than the choice of test. The accuracy of MMSE and GPCOG displayed very similar curves when plotted on ROC curves in Figures 8-12. Whilst there was no clear superiority in threshold across MMSE <24, <25 and <27, with the five included studies the sample may be too small, or with variability due to other test factors, to draw firm conclusions. Two of the studies[119,142] had uneven sample sizes which may have adversely affected variability. The challenge comes when choice of threshold is influenced by other factors such as professional guidelines, for example the optimum threshold is at odds with policy body standards, or clinical relevance, where for example the optimum threshold for a test does not sit within the observed clinical range of the patient group.

### 5.4.3. Quality issues

Substantial variability was observed in sensitivity and specificity across studies. Combined with assessments of study quality, risk of bias and applicability, this finding strongly suggests that the quality of included studies may be insufficient to make a fair comparitive valuation of the diagnostic accuracy of MMSE and GPCOG when used to identify dementia within a general practice setting. It is this variability within and across studies that is a fundamental factor in the lack of concrete conclusions which can be drawn around the comparative diagnostic accuracy of MMSE and GPCOG within this systematic review.

Specific examples of poor study quality were found in a lack of blinding, incorporation of one of the index tests (MMSE) within at least part of the reference standard, lack of prespecified diagnostic thresholds, missing study participants unaccounted for, exclusions on the basis of co-morbidities common in general practice (such as cardiovascular disorders, hearing loss and depression), repeat testing with index tests,

lack of general practitioners used as raters, and problematic study designs. Alongside issues of study quality, problems were found in the quality of reporting where clear, simple details on replicability, validation and administration were not reported for the two established tools being assessed.

Test performance was highly variable across studies shown in the forest plot in Figure 8. These disparities were potentially due to variation in assessment process, application of the tests, administrators, raters, blinding, reference standards, or other myriad sources of systematic variation presented within the Results section. This variability was observed across a wide area, so it is unclear whether one of these factors had greater influence on test performance than others or whether a combination of variables affected overall test performance.

It may be that the tools themselves are not fit for purpose, and therefore these other variables are simply chaff masking the 'true' accuracy of the brief cognitive assessments. It is not possible to know this for sure until evidence can be accurately compared across studies with confounding factors accounted for and minimised as far as possible. This set of circumstances poses some methodological challenges – some general to evidence synthesis, and some specific to diagnostic accuracy – which are addressed further in Chapter 6 of this thesis.

Table 20 Summary of results from included studies

| Study | RS | GPCOG Patient | | | | GPCOG Informant | | | | GPCOG Total/combined | | | | GPCOG Two-stage | | | | MMSE <27 | | | | MMSE <25 | | | | MMSE <24 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | Sn % | Sp % | AUC | N | Sn % | Sp % | AUC | N | Sn % | Sp % | AUC | N | Sn % | Sp % | AUC | N | Sn % | Sp % | AUC | N | Sn % | Sp % | AUC | N | Sn % | Sp % | AUC |
| Basic 2009[287] | 1,2,3,4 | 151 | 98 | 67 | 0.95 | 151 | 94 | 49 | 0.97 | 151 | 98 | 77 | 0.97 | | | | | | | | | | | | | 151 | 90 | 88 | 0.93 |
| Brodaty 2002[142] | 5:1,3,6 | 282 | 82 | 70 | 0.86 | 202 | 89 | 66 | 0.86 | 202 | 82 | 83 | 0.91 | 246 | 85 | 86 | 0.91 | | | | | 283 | 81 | 76 | 0.91 | | | | |
| Brodaty 2016[119ǂ] | 1,3,6 | | | | | | | | | 1717 | 79 | 92 | 0.92 | 1717 | 80 | 91 | NR | | | | | | | | | 2028 | 51 | 97 | 0.91 |
| Li 2013[†288] | 1,7 | 356 | 97 | 66 | 0.97 | 200 | 99 | 82 | 0.96 | 356 | 97 | 86 | 0.97 | 356 | 97 | 89 | 0.97 | | | | | | | | | 356 | 87 | 94 | 0.97 |
| Pirani 2010[‡286] | 1,3,6,7,8,9 | 200 | 98 | 54 | 0.96 | 200 | 83 | 75 | 0.86 | 200 | 92 | 88 | 0.96 | 200 | 82 | 92 | 0.96 | 200 | 93 | 91 | 0.96 | 200 | 78 | 98 | 0.96 | | | | |

N, number in study population; Sn, sensitivity; Sp, specificity; AUC, area under the curve; NR, not reported ["binary data"]; RS, reference standard
†GPCOG-C (General Practitioner Assessment of Cognition – Chinese version)
‡ GPCOG-It (General Practitioner Assessment of Cognition – Italian version
ǂ Data based on different sample sizes, 313 participants missing between MMSE and GPCOG assessments

Reference standards:
1 DSM-IV (Diagnostic and Statistical Manual version 5)
2 CDR (Clinical Dementia Rating)
3MMSE (Mini Mental State Examination)
4GPCOG (General Practitioner Assessment of Cognition)
5CAMDEX (Cambridge Examination for Mental Disorders of the Elderly Examination)
6CAMCOG (Cambridge Examination for Mental Disorders of the Elderly Cognitive Scale-Revised)
7clinical judgment (of attending clinician)
8ADAS-Cog (Alzheimer's Disease Assessment Scale Cognitive Subscale)
9physical exam

## 5.5. Discussion

A summary table of results of included studies is presented in Table 20. It is clear from this table that there is a great deal of variation across the five included studies. Inconsistency was not only observed at the test level. Both within and across included studies, substantial variation was observed in accuracy, sampling, test administration, rating, application of reference standard and analysis as described within the quality issues above. This is a suprising finding, given the comparative nature of the included studies identified according to clear prespecified criteria published in PROSPERO and within an open access peer-reviewed journal. Test performance varied widely across both MMSE and GPCOG when directly compared to one another, as well as within test categories when viewing MMSE accuracy and GPCOG accuracy individually.

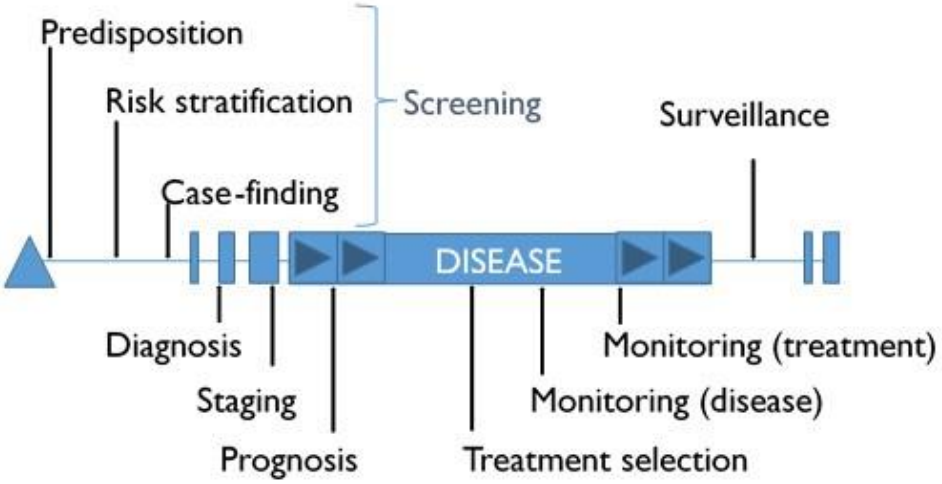
Whilst it is important to establish how good a test is at discriminating between people with and without a condition, it is equally important to recognise accuracy as part of a bigger process for evaluating a medical test. The ultimate value of a test is in the impact it has on the patient, and this cannot solely be measured through diagnostic accuracy[308]. As part of the overview and systematic review processes described in previous chapters, the data were revisited to see what information was available on outcomes beyond diagnostic accuracy. A substantial body of work was found on test administration time (how long a test takes to conduct), and some evidence on aspects of clinical effectiveness such as face validity and bias. Some evidence was also found of the broader impacts of using brief cognitive assessments.

### 5.5.1. Analytical and clinical performance measures

There was no analytical performance evidence presented within the five studies included in the systematic review of comparative diagnostic accuracy (see Table 21).

Clinical performance is the ability of a test to distinguish patients with a particular clinical condition or physiological state. Methods of distinguishing patients in this way could include assessing the accuracy of a diagnostic test, assessment of prognostic factors that predict a certain outcome, or carrying out a test that monitors a disease state or treatment outcomes (see Figure 16).

Figure 16 Some purposes of medical tests at different stages of the clinical pathway




Adapted with permission from Patrick Bossuyt.

As these issues deal with validating the test in a clinical sample and setting, this can be termed 'clinical validity' and includes features of test performance or operation, and the various measures of efficacy of a test[309], including the diagnostic or prognostic accuracy.

As this thesis focuses explicitly on the diagnostic accuracy of brief cognitive assessments, it would be expected that clinical performance measures should feature strongly throughout the evidence from both the overview and systematic review of direct comparisons.

In reality, there was scant evidence of clinical performance beyond diagnostic accuracy. Inter-rater and test-retest reliability were reported for GPCOG, GPCOG-It and GPCOG-C in three studies[142,286,288] and inter-rater reliability for MMSE in one study[288] as shown in Table 21.

### 5.5.2. Cost effectiveness evidence

There was no evidence on cost effectiveness found within the systematic review. This finding is expected given that studies of diagnostic accuracy rarely include health economic data, and these assessments are often conducted as stand-alone studies of cost effectiveness.

Data on the cost effectiveness of dementia diagnosis does exist and is widely available for primary care and community settings[310-312] but these data were not presented either within the overview or systematic review.

### 5.5.3. Broader impact evidence

Within the context of assessing brief cognitive assessments as part of the process for identifying dementia, there is much resonance between the patient focus of measures of clinical utility, and the wider assessment of the impact on patients including emotional, social, legal and ethical effects[313]. An example of these patient-centred and broader societal effects of testing is the commonly-quoted psychological effect of earlier dementia diagnosis where brief cognitive assessment contributes to the reassurance that a diagnosis may offer a patient and their carers. This benefit is not, however, guaranteed and both patients and clinicians report valuing the "right time" over early dementia diagnosis [7,314]. Equally, the ethical and legal implications of a diagnosis of dementia on individual autonomy in driving are well researched[315-317].

Timings were reported in a variety of ways from exact timings to mean averages, ranges and 'less than' statements and these issues have been discussed in some detail above.

Table 21 Systematic review-identified evidence from studies of factors beyond diagnostic test accuracy in direct comparisons of GPCOG and MMSE

| **Studies** | | **Basic 2009** | | **Brodaty 2002** | | **Brodaty 2016** | | **Pirani 2010** | | **Li 2013** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Index tests** | | GPCOG | MMSE | GPCOG | MMSE | GPCOG | MMSE | GPCOG-It | MMSE | GPCOG-C | MMSE |
| **Clinical performance** | Inter-rater reliability | | | X | | | | X | | X | X |
| | Test-retest reliability | | | X | | | | X | | X | |
| **Clinical effectiveness** | Face validity | | | | | | | | | | |
| | Internal consistency | | | X | | | | | | X | |
| | Education bias | X | X | | | | | X | | | |
| | Language/cultural bias | X | X | | | | | X | | | |
| **Broader impact** | Time taken (mins) | | | *[1] | | 4-6 | 10 | 4.45[*2] | 8.9[*2] | 4.3[*3] | 6.5[*3] |
| | Ease of admin | | | X | | | X | | | | |
| | Practicability for general practice | | | X | | | X | | | | |
| | Clinician preference | | | X | | | | | | | |
| | Acceptable to patients | | | X | | | | | | | |
| | Economically viable | | | X | | | | | | | |

x , Reported/measured; blank, not reported/ measured

*[1] 3.3 mins for patient section (±1.08, range 2-5.8 minutes); 1.23 mins for informant section (±0.64, range 0.5-2.5 minutes)

*[2] 4.45 for GPCOG-It total (±0.42); 8.9 for MMSE (±1.1)

*[3] 4.3 for GPCOG-C total (± 2.4); 6.5 for MMSE (±2.1) $p < 0.05$

In addition to extensive variability within this systematic review, various challenges have been encountered. Some of these are specific to conducting systematic reviews of comparative accuracy studies, and some are problems applicable to all accuracy studies. These issues are now addressed in detail.

### 5.5.4. Issues specific to systematic reviews of comparative accuracy

Of particular relevance to systematic reviews of comparative accuracy studies were the questions presented by the use of summary ROC plots. It was possible to pair the two tests for each study, as shown in Figure 9 and Figure 10 – but beyond observed variability, particularly clear in Figure 10, this was not greatly additive to the information presented by study authors within individual ROC curves. More generally, there may be a challenge in the topic of study when considering how useful ROC curves are for further analysis of the data. Within psychometric testing, the benefit of assessing ROC curves may be more clear (where thresholds are often more readily tailored to the population and setting under assessment) than in other areas of interest where – for example – cut points are restricted or tightly defined in practice.

*Incorporation bias*

Whilst incorporation bias may be more commonly observed in comparative diagnostic accuracy studies due to the greater number of index tests and greater complexity in study design, the risk of bias introduced by incorporating an index test into the reference standard is not well understood[305]. It is often assumed that incorporation bias would risk overestimating the size of effect of an index test due to artificially close agreement with the reference standard[318]. Alternatively, in some cases, incorporation bias may provide a disadvantage in the extent to which asymmetry is observed. Inadequacies of the index test when used in practice may lead to system changes which mean when the test is used subsequently as an index test it performs less well compared to another test not used in practice. These are scenarios worth considering for the impact of incorporating index tests as reference standards on pragmatic test performance, without fully exploring the testing pathway.

*Study design*

The underlying premise of comparative accuracy studies is that by excluding possible confounding variables introduced through indirect evaluations, such as deviations in populations, recruitment and sampling, comparisons made are more equal and differences more likely due to actual variation in accuracy. This does depend in part

on study design, with common designs including cross-sectional and case-control designs. In cross-sectional designs, participants are recruited as a single sample and stratified into cases and non-cases using the reference standard, with the performance of the index test assessed against the performance of the reference standard[301]. With case-control designs, a sample of known cases are compared to a different sample of controls. This may introduce variation between sample populations, either as an artefact of the sampling or some unknown quality of the selected populations. These differences may not be known, and yet still present a source of variation within the study. Whilst cross-sectional designs draw on a single sample so potential bias would theoretically affect both arms equally, case control study designs have the potential to introduce asymmetric variation due to systematic differences at the population and/or sampling levels[302]. Systemic variation of this type becomes more difficult to anticipate or account for when different study designs are combined as was the case in this systematic review.

In practical terms, case control designs will feature in systematic reviews unless excluded as part of the selection criteria (which is not recommended within the Cochrane Handbook for Diagnostic Test Accuracy[301]), and so systematic reviews of comparative accuracy should stratify analysis by study design. How this is methodologically handled in a systematic manner is yet to be established.

*Uneven sample sizes*

Brodaty 2002[142] used a cross-sectional study design of 283 participants with mixed symptomatic and asymptomatic presentation. Asymptomatic patients were aged 75 years and over, and symptomatic patients were aged between 50 and 74 years old with a 'suspected' memory problem. In this study there were discrepancies amongst all the sample sizes presented, and different numbers reported across all measures (GPCOG Patient: N=282; GPCOG Two Stage: N=246; MMSE: N=28; AMT: N=269; GPCOG Informant and GPCOG Total: N=202). Eighty-one participants went missing between recruitment and administration of the MMSE to conduct of the GPCOG Total, although without a study flow diagram provided it is unclear at what stage the participants went missing. This constitutes a loss of 29% of participants (283/202) and this loss is distributed asymmetrically within the GPCOG sample (i.e. all the missing data relates to GPCOG Total, whereas the total sample is reflected in the MMSE data).

The study authors noted that sample sizes varied due to missing data but made no further reference to the data [142].

In a later single-gate cross-sectional study, Brodaty 2016[119] recruited 2028 participants living in the community with mixed symptomatic and asymptomatic presentation. The exact mix of symptoms and allocation within the study is unclear from the manuscript. All participants did not complete all the brief cognitive assessment measures; 2028 participants were administered the MMSE, but only 1717 undertook GPCOG with missing GPCOG data of 311 participants unaccounted for. This loss represents 15% of all participants within the study (1717/2828) and as with Brodaty 2002 the distribution of missing data is asymmetric, as only the GPCOG data is missing.

This missing data is recognised within the QUADAS-2 quality assessment, but was not immediately obvious within the research papers – partly due to several sub-tests being administered, and partly due to a lack of PRISMA-type flow diagrams presented. The reality was that there was a need to explore the 2x2 table before realising there were flaws in the detail of the study, including missing data. This led to a dilemma of conducting the systematic review; whilst these studies had been included within the review, was the potential bias introduced by this unaccounted-for missing data so severe that these studies should be excluded on this basis? The primary author on these two studies was contacted to ask for reasons for this imbalance and whilst he confirmed he would investigate the data, and a response has not yet been received. It was decided to treat this missing data as a major limitation for these studies, but rather than excluding them from analysis to acknowledge this limitation throughout and draw inferences with additional caution, bearing in mind both the missing data and asymmetry in the potential bias this may introduce.

*QUADAS-2 quality assessment tool*

The assessment of quality was carried out with the help of the QUADAS-2 tool. This was designed to aid quality assessment of diagnostic accuracy studies for inclusion in systematic reviews, and was well suited to the task. One limitation of this tool is that it was not specifically designed for use in systematic reviews of direct comparisons of diagnostic test accuracy, where more than one index test will be present. One of the foundational beliefs in assessing diagnostic accuracy across direct comparisons is that this direct comparison allows the researcher to minimise all confounding effects

normally observed in the comparison of indirect studies (such as differences in population, setting and administration) so that the comparisons of accuracy are as fair and equal as they can be. If this is accepted, then variations in sampling (e.g. using two gate designs) and study design (such as case-control) become far more important markers of study quality[302] and need to be addressed specifically and in a more sophisticated manner within appropriate quality assessment tools.

The current version of QUADAS-2 lacks specific questions on comparative accuracy where more than one index test is present, incorporation of index tests into reference standards and assessment of factors beyond sensitivity and specificity. The quality of primary comparative accuracy studies presented challenges where in four out of five studies the reference standard was at least partly formed by one of the index tests (MMSE), two studies [119,142] had a substantial volume of missing data, two studies[286,288] used a case control study design which is known to exaggerate diagnostic accuracy[305], and there was poor reporting of key information such as test thresholds.

A modified version of QUADAS-2 making it more relevant to comparative diagnostic accuracy studies has been suggested by Wade and colleagues[319], with specific additional questions relating to clinical practice (was the execution of both tests as they should be performed in clinical practice?), independence between tests (were the index and comparator test independent? Were the reference standard and the index/comparator tests independent of each other?), verification (were the results of both tests verified using the same reference standard?), consistency (did the whole sample undergo both tests?), and missing data (was there a difference in the number of uninterpretable or indeterminate results between tests, which is likely to have biased the study results?). Incorporating these items into a modified version of the QUADAS-2 suitable for comparative accuracy studies would be an invaluable starting point to improve the methodological and qualitative evaluation of this design.

### 5.5.5. Issues applicable to all accuracy studies

Whilst challenges observed so far have been more applicable to comparative accuracy reviews, there are other issues common across all accuracy studies. It is common for studies of diagnostic accuracy to report sensitivity and specificity figures, and in some studies authors report actual numbers of patients distributed within the 2x2 binary classification table. It is unusual for study authors to explore the nature of misclassification as well as the actual numbers of false positives and false negatives,

yet this would be valuable information when comparing two or more index tests, even as indirect comparisons.

Understanding why individual patient scores differ between index tests may also help to identify key factors of the test or testing process which affect test performance. An extension of this idea would be the potential offered in being able to identify individual patient-level data where, for example, a participant's scores were high on one test and low on another. Arguably even more useful would be the ability to identify where all index tests correctly identified the same person, and whether there were any characteristics of the tests or the participant which made this concordance more or less likely.

Common problems of study quality were found where study participants did not match the target general practice population, but were instead designated based on symptoms, recruited from specialist clinics and selected for synthetic factors without credibility for the target population (such as having an informant available). In some cases, test administrators were specialised research nurses, geriatricians and psychogerontologists and did not correspond to the intended general practitioner user. Potential participants were excluded on the basis of common general practice comorbidities and potential confounders (e.g. delirium, cardiovascular conditions, hearing impairment), and there was a lack of blinding across a number of studies to at least one of the index tests.

The general quality of reporting varied substantially, and in some cases index test thresholds were not prespecified - or specified at all. Finally, on other factors that could affect test performance, only three studies[142,286,288] reported test administration time with limited detail, and one study[142] reported summarised findings of a GP satisfaction survey and a patient five-point attitude scale (findings of the patient attitude scale were not reported within the included paper).

Study designs were also varied and posed particular challenges. Three studies[119,142,287] employed two-gate, cross-sectional designs with either symptomatic[287] or mixed symptomatic and non-symptomatic[119,142] populations. Whilst mixed presentation populations may mean findings are better suited for general practice, a study population presenting with known symptoms is already a different group of people than one may expect to encounter in general practice. Perhaps of

more concern, two studies used a case-control single-gate study design. This meant that the study population was already problematic for generalisability due to the selection of the participants for cases or controls unrelated to the wider setting. Also for this reason, prevalence calculations are inappropriate with case control studies as they cannot fairly be related to a particular setting.

A key conclusion is therefore that comparative accuracy studies do not guarantee optimum study quality or superiority over non-comparative diagnostic accuracy studies. There are many methodological issues that need to be addressed in order to realise the potential that comparative studies offer to improve the evidence base of tests.

## 5.6. Recommendations

A major challenge within this systematic review is that the brief cognitive assessment tools are designed to measure cognitive function, and in many cases only some aspects of cognitive function. Whilst dementia as an umbrella term incorporates cognitive decline, there are many other aspects of dementia that are not assessed using brief cognitive assessment tools. The stated aim within included studies of identifying dementia primarily using these tools is therefore highly problematic and misrepresentative both for research purposes and for clinical practice. It is recommended that researchers moderate their language around identifying dementia, and instead specify the particular aspects of dementia diagnosis being addressed within the specific study.

For evidence based research and practice, we must ask whether the questions posed within these studies and the wider research arena are actually addressing the issues that general practitioners and patients are interested in. There is a growing need to address functional issues related to the dementia syndrome, rather than to bolster uninformative and misleading labels based upon research requirements at odds with the clinical reality of patients and GPs. Particularly at the level of primary care, taking a functional approach means focussing on individual patient needs and abilities rather than applying a catch-all term with little consequent post-diagnostic support. This would seem a more progressive approach to research with direct clinical application at its centre.

Addressing research methods, the use of case-control designs in diagnostic accuracy studies raises questions about the presence of spectrum bias, alternative conditions and comorbidities. When considering spectrum bias, it is important to recognise that the ability of a test to identify a target condition depends on the severity of the condition [320]. Dementia is a good example of this situation. As severity of dementia increases so symptoms become more marked, and thus more readily-identified with even relatively crude measures[321]. Within this systematic review, the brief cognitive assessments under investigation (MMSE and GPCOG) assess cognitive impairment. This is highly likely to be present in dementia, but is also present in many other conditions and therefore alternative diagnoses are a real possibility, leading to an increased risk of false positive results. Finally, in the general practice setting where target populations are over 50 years old, there is an increased likelihood of comorbid conditions[322,323] which in turn also elevates the risk of false positive results[302] although as many potential comorbid conditions were excluded within included studies this obfuscates actual diagnostic performance yet further.

Incorporation bias was a substantial concern within the studies included in this review, as the MMSE formed some or part of the reference standard in four out of five included studies leading to increased potential for overestimating the accuracy of MMSE as an index test. Substantial uneven sample sizes of 15%[119] and 29%[142] were unaccounted for, both within the study manuscripts and in following up with the study author. The asymmetric distribution of these missing data solely within the GPCOG arm meant that any bias introduced would not have affected both index tests equally, which was an additional cause for concern.

Within the limited evidence provided by a small number of included studies, the inference is that GPCOG offers similar performance to MMSE. Decisions need to be made on which threshold gives the optimum balance between sensitivity and specificity for referrals by GPs to secondary care. Arguably a threshold which minimises errors would be reasonable, as it is difficult to decide whether the disutility of false positive (FP) results or false negative (FN) results is greater without further investigation beyond the scope of this review. Some information on the severity of disease missed in FNs, and the extent of additional investigation required as a result of FPs would be helpful to inform this level of clinically-relevant decision making.

Considering the problems with the evidence identified within this systematic review, this raises the question of how to handle challenges to validity that are so substantial the review author might consider excluding the evidence? In this current example, there were three major concerns in a) study design; b) missing data; and c) incorporation bias. These factors could all have been eradicated by good study design and conduct. It is this simple – by avoiding case-control designs, ensuring even sample sizes distributed equally across both/all testing arms, and using a reference standard independent of the index tests (as appeared to be the case in one study[288]) these issues would be minimised or eliminated. Where case-control designs cannot be avoided, systematic reviews can explicitly either exclude other study designs or stratify synthesis so that these study designs are kept separate from others, with implications drawn relevant to that design and associated conduct. The major improvement is that studies should be well-conducted and well-reported.

There would be significant value in a well-conducted systematic review of comparative diagnostic accuracy for MMSE and GPCOG including test performance factors such as administration time, acceptability for clinicians and patients, and cost-effectiveness. Additional value may be added by taking a three-stage approach to analysis, firstly assessing the evidence according to the number of arms there are to the study, i.e. single-gate, double-gate; assessing differences between arms of individual studies; and finally assessing any overall difference. This would allow disaggregation of variation observed within and across different arms of different studies, which could highlight actual differences between tests set against variation due to other factors.

This systematic review of comparative accuracy studies has further highlighted the importance of well-conducted studies of direct comparisons, and the need to better understand factors influencing diagnostic performance beyond accuracy measures of sensitivity and specificity. It is these aspects of a test's performance beyond diagnostic accuracy that will be explored further in the next chapter.

# Appendix 8. Review protocol

Hunt *et al. Diagnostic and Prognostic Research* (2017) 1:14
DOI 10.1186/s41512-017-0014-1

Diagnostic and Prognostic Research

PROTOCOL Open Access

## The comparative diagnostic accuracy of the Mini Mental State Examination (MMSE) and the General Practitioner assessment of Cognition (GPCOG) for identifying dementia in primary care: a systematic review protocol

Harriet A. Hunt[1*], Sanne Van Kampen[2], Yemisi Takwoingi[3], David J. Llewellyn[4], Mark Pearson[1] and Christopher J. Hyde[5]

**Abstract**

**Background:** Improved dementia identification is a global health priority, and general practitioners (GPs) are often the first point of contact for people with concerns about their cognition. However, GPs often express uncertainty in using assessment tools and the evidence based on which tests are most accurate in identifying dementia is unclear. In particular, there is little certainty around how the accuracy of available brief cognitive assessments compares within a clinical family practice setting.
Grounded in existing brief cognitive assessment evidence, we will compare the diagnostic test accuracy of the Mini Mental State Examination (MMSE) to the General Practitioner Assessment of Cognition (GPCOG) against the best available reference standard when used within a family practice setting.

**Methods:** We will employ robust systematic review methods to assess studies of diagnostic accuracy where both the MMSE and GPCOG have been evaluated as direct comparisons, i.e. within the same study population. This approach will enable us to minimise between-study heterogeneity, to eliminate the risk of bias due to confounding and increase the opportunity to make clinically useful and useable comparisons of diagnostic accuracy across both the MMSE and GPCOG. This systematic review will be conducted using a pragmatic search strategy, refining searches that build upon studies identified as part of our overview of systematic reviews of the diagnostic accuracy of brief cognitive assessments for identifying dementia in primary care.

**Discussion:** Through this systematic review, we aim to improve existing evidence on how the diagnostic accuracy of MMSE and GPCOG compares when used to identify dementia within the family practice setting. We also aim to make clinical practice recommendations based upon the variations in diagnostic accuracy identified between the MMSE and GPCOG.

**Keywords:** Dementia, Brief cognitive assessment, Diagnostic accuracy, Primary care, Systematic review

* Correspondence: h.a.hunt@exeter.ac.uk
[1]National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula, Institute of Health Research, University of Exeter Medical School, St Luke's Campus, Exeter, Devon EX1 1TE, UK
Full list of author information is available at the end of the article

## Background

Improved dementia diagnosis is a global health priority of international bodies such as the World Health Organization [1] and the G8 [2]. Existing delays in the diagnostic pathway have led to debates around case finding and targeted screening within primary care [3–6]. General practitioners (GPs) are often the first point of contact for people with concerns about their cognition, yet GPs often express uncertainty in using assessment tools alongside concerns around the consequences of misdiagnosing dementia [5, 7, 8]. In established healthcare systems, guidelines on the most accurate brief cognitive assessment for identifying dementia in primary care are inconsistent and variable in their specific recommendations. Whilst there is variation in guidance on thresholds, accuracy and suitability of test within different populations, guidelines often feature the same subset of brief cognitive assessments. Examples are available in the UK from the National Institute for Health and Care Excellence [9] and the Royal College of Psychiatrists [10], and in the Netherlands from the Huisartsen Genootschap (GP Society) [11]. These all include the Mini Mental State Examination (MMSE) and the General Practitioner Assessment of Cognition (GPCOG).

A number of systematic reviews [12–18] have explored the individual diagnostic accuracy of brief cognitive assessments for dementia in isolation, and across a range of populations and settings. In an overview of systematic reviews of the diagnostic accuracy of brief cognitive assessments for identifying dementia in primary care, we identified two brief cognitive assessments that can be compared to identify the test with better diagnostic performance. These tests were two of the three most frequently assessed brief cognitive assessments within the 13 included systematic reviews [12, 13, 19–29] with the MMSE featuring in 8 reviews and the GPCOG featuring in 4 reviews. The clock drawing test (CDT) was the third most frequently assessed tool, featuring in 4 reviews. We judged this to be less comparable to the MMSE in terms of administration complexity, timing and domains assessed, compared to the GPCOG. As the most frequently assessed test within our overview, the MMSE is also included as one of the index tests within this review as, whilst copyright restrictions are now enforced, it remains one of the most popular brief cognitive assessments employed in practice [27, 30]. The MMSE is based on a 30-point scale of 11 questions testing five domains of cognitive function (orientation, registration, attention and calculation, recall and language) [31]. The GPCOG was the second most frequently assessed index test within our overview. The GPCOG is a publicly available test that has two sections: a patient examination (GPCOG-patient) with a maximum score of 9 (optimum performance) covering time orientation, clock drawing, reporting recent events and a word-recall task, and an optional informant questionnaire (GPCOG-informant) with a maximum score of 6 with questions assessing the patient's memory of recent events and their executive function [32]. In comparison, the CDT is a standard assessment where the patient is asked to draw a clock face marking the hours and then draw the hour and minute hands to correctly indicate a specific time (e.g. quarter past 3). There are a number of scoring approaches, but the Shulman method uses a 6-point scoring system [33] whilst the Sunderland method uses a 10-point system [34]. Taking into consideration the ubiquity of the index tests, the comparability of the tests mentioned above and their common use within guidelines, we have chosen to compare the MMSE against the GPCOG as index tests within this systematic review. Therefore, the aim of this systematic review is to compare the diagnostic accuracy evidence of the MMSE and the GPCOG for identifying dementia, particularly within a primary care setting and using direct (within study) comparisons.

This use of direct comparisons should reduce between-study heterogeneity and allow us to draw firm conclusions about the comparative accuracy of these brief cognitive assessments within the same or similar populations [35, 36]. To our knowledge, this type of systematic review has not previously been conducted to compare the accuracy of brief cognitive assessments for identifying dementia.

This evidence will contribute strongly to clinical practice and policy making by demonstrating the presence or absence of superiority in the diagnostic accuracy of GPCOG relative to that of MMSE for identifying dementia in primary care.

## Methods

The primary outcome is the comparative accuracy of the two tests assessed via direct comparisons, i.e. the diagnostic accuracy of the two tests are compared within the same population in a study (comparative study).

The secondary outcome of the review is to identify other common test-related factors identified by included studies, such as ease of administration or administration time. Whilst beyond our primary focus of test accuracy, these other factors may contribute to the overall usefulness of the tests when applied in a primary care setting, and we will incorporate them in our findings in order to make useful research and clinical recommendations.

This systematic review will be conducted using a pragmatic search strategy, refining searches that build upon studies identified as part of our overview of systematic reviews of the diagnostic accuracy of brief cognitive assessments for identifying dementia in primary care. Further details are given below (PROSPERO reference 42015022078).

### Overview search methods

To build the search database for the overview of systematic reviews of the diagnostic accuracy of brief cognitive assessments for identifying dementia in primary care, we searched the Cochrane Database of Systematic Reviews, EMBASE, MEDLINE and PsychINFO for systematic reviews from inception until August 2015. Search strategies are shown in the Additional file 1. According to best searching practice for diagnostic accuracy reviews, we applied no date or language restrictions, and where reviews were updated, we used the latest version available. Additional papers were identified through Zetoc alerts and incorporated at the title and abstract screening phase. We ran updated searches on the Cochrane Database of Systematic Reviews in February 2016.

#### *Eligibility criteria*

Adults aged 18 years or over recruited from a primary care or general practice population were included, and we did not exclude patients who were selected on the basis of an existing diagnosis or condition which might reasonably be expected to feature in primary care (e.g. stroke).

The target condition was all-cause (non-differentiated) dementia. We also included reviews that focused specifically on differentiated forms of dementia such as Alzheimer's disease, vascular dementia and dementia with Lewy bodies. We excluded reviews that focused on mild cognitive impairment (MCI). Where reviews investigated both dementia and MCI, we extracted data referring to dementia and excluded data referring solely to MCI.

### Identification of studies for this systematic review

To identify eligible studies for this systematic review, we will first assess the 13 systematic reviews included within our overview review (methods described above) and identify included reviews that contained studies including direct comparisons of the diagnostic accuracy of MMSE and GPCOG for identifying dementia in primary care. Once we have identified these studies, we will carry out citation tracking via Google Scholar, i.e. clicking on the appropriate link (e.g. "cited by 15") to view details on the articles that have cited the original study. We will also use these initial studies to conduct snowball searching, i.e. checking the bibliographies for relevant original studies for possible inclusion within this systematic review. We will use Zetoc alerts to proactively identify recent studies published that meet our criteria (using the terms "MMSE", "GPCOG", "test accuracy" and "dementia"). Finally, when we have identified studies using the above methods, we will conduct a traditional search taking a start date 1 year prior to the most recently published identified study up to the current day, using MEDLINE, EMBASE and PsychINFO databases. The rationale is that this search will cover the maximum period of time not covered in the overview review searches with some date overlap to ensure all potential sources are included, using the most efficient means to identify the most recent evidence. This will also enable us to confirm whether we identified all relevant studies via the overview searches.

### Index tests

The index tests are the MMSE [31] and the GPCOG [32]. The MMSE is one of the most widely used brief cognitive assessments currently used, and development of the GPCOG has been independent to the development of the MMSE.

The conventional threshold for the MMSE is 24 (also shown as <24), where out of a maximum possible 30 points, scores below 24 indicate impairment [22]. The GPCOG comprises of two sections: the section completed by the individual being assessed, known as GPCOG-patient, and an optional section for a relative or friend to complete (if present) known as GPCOG-informant. GPCOG-patient has 9 items with possible total scores of between 0 (indicating severe impairment) and 9 (indicating no impairment). GPCOG-informant has 6 items with possible total scores of between 0 (indicating severe impairment) and 6 (indicating no impairment). GPCOG-patient can be conducted by itself, with a conventional threshold of 8 out of 9 (<8). If informants are available, a score of GPCOG-patient between 5 and 8 precipitates the GPCOG-informant and the scores are combined ("GPCOG-total") with a conventional threshold of 11 out of a maximum 15 (<11). If no informant is available, the conventional threshold of 8 stands. It is also possible to conduct a staged GPCOG assessment where GPCOG-informant is only required if GPCOG-patient is scored between 5 and 8 out of 9. This is known as "GPCOG Two stage".

For our assessment, we will stratify GPCOG into 3 types of test: GPCOG-patient with a threshold of <8, GPCOG-total with a threshold of <11 and GPCOG Two stage [37].

### Reference standard

There is currently no gold standard test for identifying dementia in primary care. We will accept reference standards consisting of the following tools alone, clinical diagnosis alone or clinical diagnosis combined with one or a combination of the following assessment tools:

- Diagnostic and Statistical Manual (DSM) III/III-R/IV/IV-R,
- Clinical Dementia Rating (CDR),
- International Classification of Diseases (ICD) 10,
- Geriatric Mental State–Automated Geriatric Examination for Computer Assisted Taxonomy (GMS-AGECAT),

- Cambridge Mental Disorders of the Elderly Examination (CAMDEX),
- International Psychogeriatric Association World Health Organization (IPA-WHO) criteria.

Reference standards are selected on the basis of many variables such as common practice within individual clinics, practitioner preference, specialisation and experience of healthcare professionals and practice managers and are subject to changes in cost and fashion. Many of the globally accepted reference standards such as the World Health Organization-supported ICD and the DSM produced by the American Psychiatric Association are updated regularly; the DSM-5 (sometimes referred to as DSM-V) was released in 2013 [38], and the ICD-11 is due for release by 2018 [39].

### Data extraction, selection and coding

All sources will be managed using the latest version of EndNote software. Two reviewers will pilot the screening for titles and abstracts on the first 15 sources, and we will write screening notes to help with title/abstract and full-text screening. Title/abstract and full-text screening will be conducted by the same two reviewers, and a third reviewer will resolve any disagreements.

A bespoke data abstraction form will be piloted by two reviewers using two included studies. Key data extracted will include characteristics of included systematic reviews (references and author details, overall goal of review, date review conducted, date published, participant details), included study details (such as authors, year of study, date of publication, country of study, outcomes reported, test timings) and general review limitations as well as components of the $2 \times 2$ table (TP, FP, TN, FN) or other accuracy data such as sensitivity, specificity and disease prevalence if raw numbers are not available. The data abstraction form will be accompanied by a briefing document explaining how it should be used. Data will be abstracted by one reviewer, spot-checked by a second, with a third reviewer acting as moderator if necessary.

### Assessment of methodological quality

We will use the QUADAS-2 [40] tool to assess methodological quality of diagnostic accuracy studies for systematic reviews. Whilst this tool is developed for studies focussing on a single index test, we will assess the suitability of using the tool for studies that focus on direct comparisons of two index tests by piloting the QUADAS-2 tool on one of the included studies. We will tailor QUADAS-2 in line with suitability in assessing quality of studies using direct comparisons, for example assessing the reference standard against MMSE and then the reference standard against GPCOG.

### Data synthesis and analysis

Study of specific estimates of the sensitivity and specificity (and their 95% confidence intervals) of GPCOG and MMSE will be presented graphically on a forest plot. We will also use these forest plots and summary receiver operating characteristic (SROC) plots to visually explore heterogeneity.

We will consider possible sub-group analyses investigating, for example tests using lower and higher thresholds. Other aspects that may be suitable for investigating through sub-grouping could include variations in population details such as prevalence, and variations in cases and control groups (e.g. confirmed dementia, probably dementia, people with memory problems, healthy people).

We will perform meta-analysis if the quantity and nature of the included studies permit. Again, if data allow, we will use a hierarchical meta-regression model with test type as a covariate to estimate and compare SROC curves or summary points [36]. A priori uncertainty about thresholds for determining test positivity and the likelihood of implicit thresholds suggests estimation of SROC curves using a hierarchical SROC (HSROC) meta-regression model may be preferable [41]. However, we will consider using a bivariate meta-regression model to estimate and compare summary points [42, 43] if studies use a common threshold.

We will create a summary of result table with additional summary tables of subgroup results (potential subgroups listed above) if relevant. If feasible and appropriate, we will consider translating any summary results into natural frequencies and other metrics such as predictive values to help improve understanding by readers.

We will not assess reporting bias because its impact on diagnostic accuracy is unclear, and the tools for investigating it are in the early stages of development [44].

## Discussion

We do not foresee any practical or operational issues with the conduct of this systematic review. All differences between the protocol and systematic review will be reported in the full systematic review.

## Additional file

**Additional file 1:** Search strategy, formatted for EMBASE (OVID). (DOCX 24 kb)

### Abbreviations

CAMDEX: Cambridge Mental Disorders of the Elderly Examination; CDR: Clinical Dementia Rating; CDT: Clock drawing test; DSM: Diagnostic and Statistical Manual; EMBASE: Excerpta Medica dataBASE; FN: False negative; FP: False positive; G8: Group of Eight (consensus policy group of highly industrialised nations: France, Germany, UK, Japan, USA, Canada, Russia); GMS-AGECAT: Geriatric Mental State–Automated Geriatric Examination for Computer Assisted Taxonomy; GP: General practitioner; GPCOG: General Practitioner Assessment of Cognition; HSROC: Hierarchical summary receiver

operating characteristic; ICD: International Classification of Diseases; IPA-WHO: International Psychogeriatric Association World Health Organization; MCI: Mild cognitive impairment; MEDLINE: Medical Literature Analysis and Retrieval System Online; MMSE: Mini Mental State Examination; NHS: National Health Service; NIHR: National Institute for Health Research; PROSPERO: International prospective register of systematic reviews; PsychINFO: Bibliographic database of psychological research; QUADAS/QUADAS-2: Quality Assessment for Diagnostic Assessment Studies/2nd version; SROC: Summary receiver operating characteristic; TN: True negative; TP: True positive; WHO: World Health Organization

**Acknowledgements**
This research is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

**Funding**
No funding has been received for the conduct of this specific systematic review, although it is supported as part of doctoral research funding for HH provided by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula.

**Availability of data and materials**
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Authors' contributions**
HH created the first draft, and all authors contributed to, read and approved the final manuscript.

**Competing interests**
The authors declare that they have no competing interests.

**Consent for publication**
Not applicable.

**Ethics approval and consent to participate**
Not applicable.

## Publisher's Note



**Author details**
[1]National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula, Institute of Health Research, University of Exeter Medical School, St Luke's Campus, Exeter, Devon EX1 1TE, UK. [2]National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula, Plymouth University Peninsula Schools of Medicine and Dentistry, ITTC building, Tamar Science Park, Drake Circus, Plymouth, Devon PL4 8AA, UK. [3]Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [4]University of Exeter Medical School, St Luke's Campus, Exeter, Devon EX1 1TE, UK. [5]Institute of Health Research, University of Exeter Medical School, St Luke's Campus, Exeter, Devon EX1 1TE, UK.

Received: 10 March 2017 Accepted: 1 May 2017
Published online: 02 June 2017

**References**

1. WHO, ADI. Dementia: a public health priority. Geneva: World Health Organization; 2012. viii + 102 pp.
2. G8 U. G8 dementia summit declaration. In: Office DoHaPMs, editor. online: https://www.gov.uk/government/publications/g8-dementia-summit-agreements: Department of Health; 2013. Accessed 24 Apr 2017.
3. Dowrick A, Southern A. Dementia 2014: opportunity for change. 2014.
4. Le Couteur D, Doust J, Creasey H, Brayne C. Political drive to screen for predementia: not evidence based and ignores the harms of diagnosis. BMJ. 2013;347(7925):f5125.
5. Fox C, Lafortune L, Boustani M, Brayne C. The pros and cons of early diagnosis in dementia. Br J Gen Pract. 2013;63(612):e510-2.
6. Burns A. Alistair Burns and 51 colleagues reply to David Le Couteur and colleagues. BMJ. 2013;347(oct15_6):f6125.
7. Ballard C, Bannister C. Criteria for the diagnosis of dementia. In: David A, ABaJOB, editors. Dementia. 4th ed. London: Hodder; 2010. p. 794.
8. Robinson L, Tang E, Taylor JP. Dementia: timely diagnosis and early intervention. Bmj. 2015;350:1-6.
9. NICE Pathways. Dementia diagnosis and assessment. 2016. p. 10.
10. Barker A, Arya P, Boston P, Fawzi W, Lennon S, Mikashi S, et al. OP86: Individual patient outcome measures recommended for use in older people's mental health. London: Online: Royal College of Psychiatrists; 2012. Contract No.: OP86.
11. Moll van Charante E, Perry M Vernooij-Dassen MJFJ, Boswijk DFR, Stoffels J, L Achthoven, MN L-K. Dementia Summary Map M21 2017 Website: https://www.nhg.org/standaarden/samenvatting/dementie. Accessed 24 May 2017.
12. Arevalo-Rodriguez I, Smailagic N, Roque IFM, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev. 2015;(3)1:Cd010783.
13. Creavin ST, Wisniewski S, Noel-Storr AH, Trevelyan CM, Hampton T, Rayment D, Thom VM, Nash KJ, Elhamoui H, Milligan R, Patel AS. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. The Cochrane Library. 2016;(1):1-183.
14. Davis DH, Creavin ST, Yip JL, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. The Cochrane Library. 2015;(1):1-50.
15. Fage BA, Chan CC, Gill SS, Noel-Storr AH, Herrmann N, Smailagic N, Nikolaou V, Seitz DP. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting. The Cochrane Library. 2015;(1):1-33.
16. Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. Cochrane Database Syst Rev. 2015;(3)1:Cd010772.
17. Hendry K, Lees Rosalind A, McShane R, Noel-Storr Anna H, Stott David J, Quinn Terry J. AD-8 for diagnosis of dementia across a variety of healthcare settings. Cochrane Database Syst Rev. 2014 [cited 1 1]; (5). Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011121/abstract http://onlinelibrary.wiley.com/store/10.1002/14651858.CD011121/asset/v=1&t=idlm4c21&s=8dafa435a321ce12c678db51152a73d7cc4fcb37. Accessed 24 Apr 2017.
18. Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. The Cochrane Library. 2014;(1):1-91.
19. Brodaty H, Low L-F, Gibson L, Burns K. What is the best dementia screening instrument for general practitioners to use? Am J Geriatr Psychiatry. 2006; 14(5):391–400.
20. Carnero-Pardo C, Espejo-Martinez B, Lopez-Alcalde S, Espinosa-Garcia M, Saez-Zea C, Vilchez-Camillo R, et al. Effectiveness and costs of phototest in dementia and cognitive impairment screening. BMC Neurol. 2011;11(1):92.
21. Carnero-Pardo C, Lopez-Alcalde S, Allegri RF, Russo MJ. A systematic review and meta-analysis of the diagnostic accuracy of the Phototest for cognitive impairment and dementia. Dement Neuropsychologia. 2014;8(2):141–7.
22. Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J Psychiatr Res. 2009;43(4):411–31.
23. Mitchell AJ, Malladi S. Screening and case finding tools for the detection of dementia. Part I: evidence-based meta-analysis of multidomain tests. Am J Geriatr Psychiatry. 2010;18(9):759–82.
24. Mitchell AJ, Malladi S. Screening and case-finding tools for the detection of dementia. Part II: evidence-based meta-analysis of single-domain tests. Am J Geriatr Psychiatry. 2010;18(9):783–800.
25. Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting. The Cochrane Library. 2014;(1):1-57.
26. Naqvi RM, Haider S, Tomlinson G, Alibhai S. Cognitive assessments in multicultural populations using the Rowland Universal Dementia Assessment Scale: a systematic review and meta-analysis. Can Med Assoc J. 2015;187(5):E169–75.

27. Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: a systematic review for the U.S. preventive services task force. Ann Intern Med. 2013;159(9):601–12.
28. Tsoi KK, Chan JY, Hirai HW, Wong SY, Kwok TC. Cognitive tests to detect dementia: a systematic review and meta-analysis. JAMA internal medicine. 2015;175(9):1450-8.
29. Woodford H, George J. Cognitive assessment in the elderly: a review of clinical methods. QJM. 2007;100(8):469–84.
30. Ismail Z, Rajji TK, Shulman KI. Brief cognitive screening instruments: an update. Int J Geriatr Psychiatry. 2010;25(2):111–20.
31. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189–98.
32. Brodaty H, Pond D, Kemp NM, Luscombe G, Harding L, Berman K, et al. The GPCOG: a new screening test for dementia designed for general practice. J Am Geriatr Soc. 2002;50(3):530–4.
33. Shulman K, Shedletsky R, Silver I. The challenge of time: clock-drawing and cognitive function in the elderly. Int J Geriatr Psychiatry. 1986;1:135–40.
34. Sunderland T, Hill J, Mellow A, Lawlor B, Gundersheimer J, Newhouse P, et al. Clock drawing in Alzheimer's disease: a novel measure of dementia severity. J Am Geriatr Soc. 1989;37:725–9.
35. Takwoingi Y, Leeflang M, Deeks J. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. Ann Intern Med. 2013; 158(7):544–54.
36. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. 2010. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 10 [Internet]. Available from: http://srdta.cochrane.org/handbook-dta-reviews. Accessed 24 May 2017.
37. Brodaty H, Connors MH, Loy C, Teixeira-Pinto A, Stocks N, Gunn J, et al. Screening for dementia in primary care: a comparison of the GPCOG and the MMSE. Dement Geriatr Cogn Disord. 2016;42(5-6):323–30.
38. APA. Diagnostic and Statistical Manual of Mental Disorders (DSM–5) American Psychiatric Association website 2017 Available from: https://www.psychiatry.org/psychiatrists/practice/dsm. Accessed 24 May 2017.
39. WHO. The 11th Revision of the International Classification of Diseases (ICD-11) is due by 2018 2017 [Available from: http://www.who.int/classifications/icd/revision/en/. Accessed 24 May 2017.
40. Whiting PF, Rutjes AS, Westwood ME, et al. Quadas-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011; 155(8):529–36.
41. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001;20(19):2865–84.
42. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal and binomial-normal bivariate Summary ROC approaches. Med Decis Making. 2008;28.
43. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982–90.
44. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. Medical Decision Making. 2008;28(5): 639-49.

## 5.7. Appendix 9. Screening notes for title/abstract screening

Primary outcome: comparative accuracy of the two tests assessed via direct comparisons, i.e. the diagnostic accuracy of the two tests are compared within the same population in a study (comparative study).

Secondary outcome: to identify other common test-related factors identified by included studies, such as ease of administration or administration time. Whilst beyond our primary focus of test accuracy, these other factors may contribute to the overall usefulness of the tests when applied in a primary care setting, and we will incorporate them in our findings in order to make useful research and clinical recommendations.

### 5.7.1. Eligibility criteria

Adults aged 18 years or over recruited from a primary care or general practice population were included, and we did not exclude patients who were selected on the basis of an existing diagnosis or condition which might reasonably be expected to feature in primary care (e.g. stroke).

The target condition was all-cause (non-differentiated) dementia. We also included reviews that focused specifically on differentiated forms of dementia such as Alzheimer's disease, vascular dementia and dementia with Lewy bodies. We excluded reviews that focused on mild cognitive impairment (MCI). Where reviews investigated both dementia and MCI, we extracted data referring to dementia and excluded data referring solely to MCI.

### 5.7.2. Inclusion criteria

*Population*: Adults aged 18 years and over

*Conditions:*

- *All-cause dementia*
- *Alzheimer's disease*
- *Vascular dementia*
- *Dementia with Lewy bodies*
- *Other differentiated forms of dementia*

*Setting: Primary care or general practice*

*Index tests*: The following two index tests should both be assessed in the same population:

- Mini Mental State Examination (MMSE)
- General Practitioner assessment of Cognition (GPCOG)

*Reference standard:* reference standards consisting of the following tools alone, clinical diagnosis alone or clinical diagnosis combined with one or a combination of the following assessment tools:

- Diagnostic and Statistical Manual (DSM) III/III-R/IV/IV-R
- Clinical Dementia Rating (CDR)
- International Classification of Diseases (ICD) 10
- Geriatric Mental State – Automated Geriatric Examination for Computer Assisted Taxonomy (GMS-AGECAT)
- Cambridge Mental Disorders of the Elderly Examination (CAMDEX)
- International Psychogeriatric Association World Health Organisation (IPA-WHO) criteria

*Outcomes:* The following outcomes are relevant if they are compared between both index tests in the included study:

- Diagnostic accuracy
- Ease of test administration
- Test administration time
- Other test-related characteristics

*Study design: all (no limits)*

*Geography: worldwide (no limits)*

*Time period: all time periods (no limits)*

*Languages: English, Dutch*

## 5.8. Appendix 10. Data extraction form

| | | | | |
|---|---|---|---|---|
| Full study reference (author/date/title/journal) | | | | |
| Date study published | | Date study conducted | | |
| Study design and timing of data collection (prospective, retrospective) | | Country of study | | |
| Stated aim(s) of the study | | | | |
| Study population – age, sex (% female where given) | | Recruitment and setting (e.g. community) | | |
| Dementia severity (including assessment measures, e.g. CDR) and disease prevalence where reported | | | | |
| Timings and combinations of tests conducted (e.g. ordering, randomisation) | | | | |
| Reference standard (detail, including threshold[s] where reported, who assessed, timings of assessments etc.) | | | | |
| Index tests (list, including version if reported) | #1 | | Threshold used | |
| | #2 | | Threshold used | |
| | Etc. | | Threshold used | |
| Index test #1 accuracy data – *name of test* | Sensitivity (%) | | Specificity (%) | |
| | PPV (%) | | NPV (%) | |
| Other Index test #1 accuracy data (e.g. AUC) | | | | |
| Index test #2 accuracy data – *name of test* | Sensitivity (%) | | Specificity (%) | |
| | PPV (%) | | NPV (%) | |
| Other Index test #2 accuracy data (e.g. AUC) | | | | |
| Index test #3 accuracy data – *name of test* | Sensitivity (%) | | Specificity (%) | |
| | PPV (%) | | NPV (%) | |
| Other Index test #3 accuracy data (e.g. AUC) | | | | |
| Review limitations | | | | |

## 5.9. Appendix 11 QUADAS-2 tool

# QUADAS-2

### Phase 1: State the review question:

| |
|---|
| *Patients (setting, intended use of index test, presentation, prior testing):* |
| *Index test(s):* |
| *Reference standard and target condition:* |

### 5.9.1. Phase 2: Draw a flow diagram for the primary study

| |
|---|
| |

### 5.9.2. Phase 3: Risk of bias and applicability judgments

*QUADAS-2 is structured so that 4 key domains are each rated in terms of the risk of bias and the concern regarding applicability to the research question (as defined above). Each key domain has a set of signalling questions to help reach the judgments regarding bias and applicability.*

## DOMAIN 1: PATIENT SELECTION

### A. Risk of Bias

| Describe methods of patient selection: |
|---|

- ❖ Was a consecutive or random sample of patients enrolled? Yes/No/Unclear
- ❖ Was a case-control design avoided? Yes/No/Unclear
- ❖ Did the study avoid inappropriate exclusions? Yes/No/Unclear

**Could the selection of patients have introduced bias? RISK: LOW/HIGH/UNCLEAR**

### B. Concerns regarding applicability

| Describe included patients (prior testing, presentation, intended use of index test and setting)**:** |
|---|

**Is there concern that the included patients do not match the review question? CONCERN: LOW/HIGH/UNCLEAR**

## DOMAIN 2: INDEX TEST(S)

**If more than one index test was used, please complete for each test.**

### A. Risk of Bias

| Describe the index test and how it was conducted and interpreted: |
|---|

- ❖ Were the index test results interpreted without knowledge of the results of the reference standard? Yes/No/Unclear
- ❖ If a threshold was used, was it pre-specified? Yes/No/Unclear **Could the conduct or interpretation of the index test have introduced bias? RISK: LOW /HIGH/UNCLEAR**

### B. Concerns regarding applicability

**Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW /HIGH/UNCLEAR**

## DOMAIN 3: REFERENCE STANDARD

### Risk of Bias

Describe the reference standard and how it was conducted and interpreted:

- ❖ Is the reference standard likely to correctly classify the target condition? Yes/No/Unclear
- ❖ Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No/Unclear

**Could the reference standard, its conduct, or its interpretation have introduced bias?** **RISK: LOW /HIGH/UNCLEAR**

### Concerns regarding applicability

**Is there concern that the target condition as defined by the reference standard does not match the review question?** **CONCERN: LOW /HIGH/UNCLEAR**

## DOMAIN 4: FLOW AND TIMING

### A. Risk of Bias

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

- ❖ Was there an appropriate interval between index test(s) and reference standard? Yes/No/Unclear
- ❖ Did all patients receive a reference standard? Yes/No/Unclear
- ❖ Did patients receive the same reference standard? Yes/No/Unclear
- ❖ Were all patients included in the analysis? Yes/No/Unclear **Could the patient flow have introduced bias? RISK: LOW /HIGH/UNCLEAR**

## 5.10. Appendix 12 Tailored quality appraisal notes

### QUADAS-2 quality assessment of the evidence for the comparative diagnostic accuracy of the Mini Mental State Examination (MMSE) and the General Practitioner assessment of Cognition (GPCOG) for identifying dementia in primary care – Raters' notes

*From the protocol:*

We will use the QUADAS-2 tool to assess methodological quality of diagnostic accuracy studies for systematic reviews. While this tool is developed for studies focussing on a single index test, we will assess the suitability of using the tool for studies that focus on direct comparisons of two index tests by piloting the QUADAS-2 tool on one of the included studies. We will tailor QUADAS-2 in line with suitability in assessing quality of studies using direct comparisons, for example assessing the reference standard against MMSE, and then the reference standard against GPCOG.

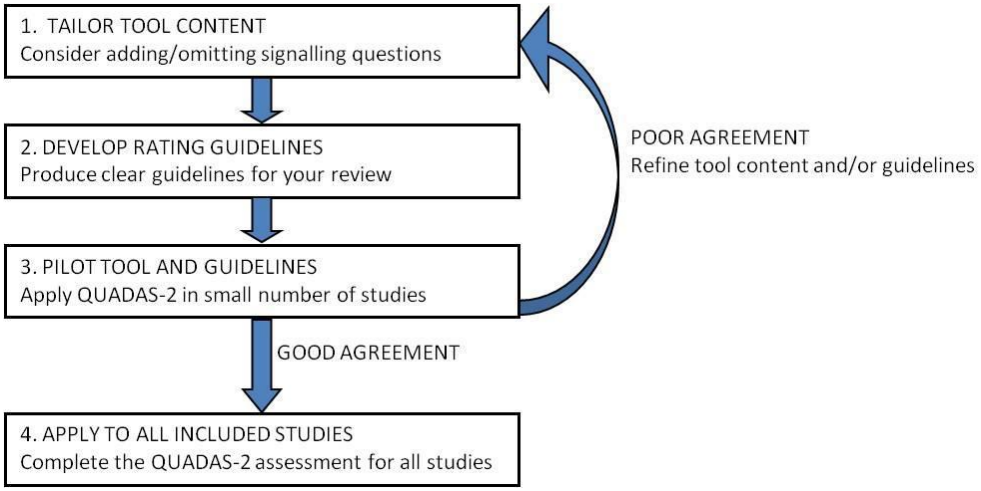
QUADAS-2 is a tool designed to assess the quality of primary diagnostic accuracy studies. QADAS-2 uses prompts (or 'signalling questions') to explore 4 main areas of potential bias and applicability: patient selection, index test, reference standard, and flow of patients through the study and timing of the index test(s) and reference standard.

*Tailoring*

We have tailored the tool to fit our particular review, which is looking at direct comparisons of two distinct index tests: MMSE and GPCOG. Therefore we have a domain assessing each index test, MMSE and GPCOG.

We will follow the QUADAS-2 advice for tailoring as shown in Figure 1.

*Figure 1 Tailoring of QUADAS-2*




*Taken from Background Notes document accessed: [http://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/](http://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/) [19/10/17]*

'Domain 2: Index tests' asks "If a threshold was used, was this pre-specified?" This was changed to simply ask "Was/were the threshold(s) pre-specified?" as both of the index tests (GPCOG and MMSE) have thresholds that should be reported.

'Domain 3: Reference standard' asks "Were the reference standard results interpreted without knowledge of the results of the index test?" This was amended to ask "Were the reference standard results interpreted without knowledge of the results of both index tests?" to include consideration of both MMSE and GPCOG.

| Glossary | |
|---|---|
| **ADI** | Alzheimer's disease International |
| **AGECAT** | Automated Geriatric Examination for Computer Assisted Taxonomy |
| **AMT(S)** | Abbreviated Mental Test Score |
| **APEx** | Exeter Collaboration for Academic Primary Care |
| **BCA** | Brief cognitive assessment |
| **BOS** | Bristol Online Surveys |
| **CAMCOG** | The Cambridge Cognitive Examination |
| **CAMDEX** | Cambridge Mental Disorders of the Elderly Examination |
| **CASI** | Cognitive Abilities Screening Instrument |
| **CCCDTD** | Canadian Consensus Conferences on the Diagnosis and Treatment of Dementia |
| **CCG** | Clinical Commissioning Group |
| **CDPC** | Cognitive Decline Partnership Centre |
| **CDR** | Clinical Dementia Rating |
| **CDT** | clock drawing test |
| **CERAD** | Consortium to Establish a Registry for Alzheimer's Disease |
| **CHAIN** | Contact, Help, Advice and Information Network |
| **CHERRIES** | Checklist for Reporting Results on Internet E-Surveys |
| **CPG** | Clinical Practice Guideline |
| **DSM-III/ III-R/ IV/ IV-R** | The Diagnostic and Statistical Manual of Mental Disorders (version 3/ version 3 revised/ version 4/ version 4 revised |
| **EMBASE** | Excerpta Medica dataBASE |
| **FAB** | Frontal Assessment Battery |
| **FAQ** | Functional Activities Questionnaire |
| **FCSRT** | Free and Cued Selective Reminding Test |
| **GMS-AGECAT** | Geriatric Mental State Schedule - Automated Geriatric Examination for Computer Assisted Taxonomy |
| **GP** | General Practitioner |
| **GPCOG** | The General Practitioner assessment of Cognition |
| **ICD-10** | International Classification of Disease – version 10 |
| **IQCODE** | Informant Questionnaire on Cognitive Decline in the Elderly |
| **IPA/WHO criteria** | International Psychogeriatric Association/World Health organisation criteria |
| **KICA-Cog** | Kimberley Indigenous Cognitive Assessment |

| | |
|---|---|
| **KICA-Screen** | Kimberley Indigenous Cognitive screening tool |
| **MCI** | Mild Cognitive Impairment |
| **MEDLINE** | Medical Literature Analysis and Retrieval System Online |
| **MeSH** | Medical Subject Headings |
| **MIS** | Memory Impairment Screen |
| **MMSE** | Mini mental state examination |
| **MoCA** | Montreal Cognitive Assessment |
| **NHS** | National Health Service (UK) |
| **NICE** | National Institute for Health and Care Excellence (UK) |
| **NINDCS-ADRDA** | National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association |
| **NINCDS-AIREN** | National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l'Enseignement en Neurosciences |
| **NINCDR-CERAD** | National Institute of Neurological Disorders - Clinical Dementia Rating -Consortium to Establish a Registry for Alzheimer's Disease |
| **PAS** | Psychogeriatric Assessment Scale |
| **PBKDF2** | A password hashing algorithm |
| **PCL** | Prueba cognitive de leganes |
| **PROSPERO** | International Prospective Register of Systematic Reviews |
| **PsychInfo** | Database of abstracts of literature in the field of psychology. |
| **RUDAS** | The Rowland Universal Dementia Assessment Scale |
| **SASSI** | Short and Sweet Screening Instrument |
| **SHA256** | A cryptographic hash function |
| **SIS** | Six item screener |
| **SPMSQ** | Short portable mental status questionnaire |
| **SSL** | Secure Sockets Layer |
| **TICS** | Telephone Interview for Cognitive Status |
| **TRIP** | Turning Research Into Practice database |
| **TYM** | Test Your Memory |
| **UK** | United Kingdom |
| **VF-an** | Verbal Fluency - animals |
| **WHO** | World Health Organisation |

# 6. The clinical reality of identifying dementia using brief cognitive assessments as part of the primary care consultation

*"Case finding" is the recourse of those who do non-evidence based screening but can't seem to admit it.*

Margaret McCartney, general practitioner in BMJ 2018; 362:k3745

This chapter presents findings from a survey of GPs exploring how general practitioners choose and use brief cognitive assessments in practice as part of the process for identifying possible dementia in primary care. It was expected that GPs would select the most suitable BCA based on best current evidence and national clinical practice guidelines and the major barrier reported would be administration time and lack of clinician understanding for how to interpret the test. The survey was carried out between June and September 2017. MMSE, GPCOG and 6CIT were the most frequently-used tests for initial assessment of cognition and for monitoring function over time. How long a test took to administer was the most popular factor influencing respondents' choice of brief cognitive assessment. On average, GPs felt that 5 minutes was an acceptable length of time for a brief cognitive assessment to take if a GP carried it out, and 10 minutes was acceptable if another member of practice staff carried out the assessment. Qualitative data revealed that major issues for GPs surveyed included the length of the assessment process, and tensions between the quality of a test and how quick or easy it was to carry out.

## 6.1. Background: why this survey was needed

Whilst the timely identification of dementia and referral to specialist services is generally acknowledged to be important to improve the chances of living well with dementia for as long as possible, general practice assessment needs to be balanced with implications of the increased likelihood of false positives and moderate test performance of brief cognitive assessments within this setting[129,156,324]. Recommendations for operationalizing the assessment of cognitive performance and detection of dementia have been created as part of national dementia strategies by several countries including the United States[325], the UK[326,] and Australia[327]. These national dementia strategies all discuss brief cognitive assessment tools in detail, as well as highlighting the value of employing appropriate tools as part of an iterative

diagnostic process. These strategies also commonly acknowledge many of the challenges in finding the most suitable tool for the primary care setting. What they lack is explicit, current and evidence-based guidance on how the best tool should be chosen by the primary care practitioner, and when tools should – and should not – be recommended for use as part of the diagnostic routine.

Despite the existence of many brief cognitive assessment tools in primary care and increasing evidence that people prefer to receive a diagnosis of dementia as soon as possible[35], current evidence is mixed for how GPs view and value the place of these brief cognitive assessments within the diagnostic process. GP attitudes to diagnosis of possible dementia in family practice have been investigated widely in recent years, commonly finding a reticence in GPs to diagnose dementia early, or to even use the word 'dementia'[328-331]. Whilst GP trainees report feeling positive about their future role in identifying people with dementia earlier[332], existing practitioners can be reluctant to either investigate dementia or disclose the diagnosis where the practitioner perceives no consequential benefit to the patient (such as access to effective therapy or additional support services)[17,49,329,333-335]. A recent survey found GPs did not tend to refer their oldest patients for further diagnostic testing unless specifically requested to do so, as they felt there was insufficient support for further treatment[49]. Even in countries where post-diagnosis support is broadly considered adequate, GPs report feelings of therapeutic nihilism and a reluctance to speak openly to patients about dementia[328]. Many GPs and GP trainees reported they had received insufficient education in dementia recognition, or would like more[330,332]. These findings may suggest motives behind continued low rates of diagnosis reported globally[336], despite national and international drives to increase diagnosis rates at earlier stages of the syndrome[326,337].

The picture, however, is unclear around how useful GPs consider brief cognitive assessment tools, how they actually choose and use these tools and which tools they favour if any. In one survey[338], when asked specifically about the effectiveness of brief cognitive assessments used most frequently in their practice, GPs rated most tools as 'good' – the mid-point on a 5 point scale from 'very poor' to 'excellent'. When respondents were asked to rank the top attributes of a cognitive screening tool for use in general practice, 'validity/accuracy' was the top element identified followed by 'ease of administration' and 'time required' [339].

Recent interviews with Dutch GPs explored their practices and views of diagnosing dementia. When questioned on their preferences for future roles in the diagnostic process for dementia, most GPs were keen for further involvement yet a number identified the inadequacy of current tests and lack of time for assessment as a barrier or concern [49]. This finding was echoed in interview research from Ireland, where GPs highlighted the complexity of interpreting the results of some brief cognitive assessment tools such as the MMSE, citing educational bias and a lack of useful thresholds as a concern [340]. Concerns around the applicability of MMSE for people with language challenges, as well as a lack of sensitivity in identifying subtle changes were also voiced in interviews with Dutch GPs [49].

Reported practitioner concerns of a lack of suitable assessment tools or a lack of tools perceived as helpful was a common finding in the literature [50,69,72,340-342], yet there is sparse evidence for how GPs actually use – or do not use - these tests within clinical practice. Recent survey evidence indicates GP uncertainty when interpreting the results of brief cognitive assessments and incorporating results to help form a management strategy [340] yet does not identify if or how this uncertainty is dealt with.

In addition to diagnostic accuracy, many other factors affect a tool's use such as how long it takes to administer [10,49,341], suitability for the target population [10,343], adaptability for different scenarios [340], clarity for clinical and patient users [10], cost [43], and possible increased value in combination with other tests [84,343]. Whilst all of these factors commonly feature in comments or side references amongst GP concerns when asked about brief cognitive assessments, they have not yet been the specific focus of survey research in this area. These elements contribute important features to help understand GP use of and attitudes towards the current available brief cognitive assessments available to help identify dementia in primary care.

The aim of this survey was therefore to investigate current clinical practice in the use of brief cognitive assessments to evaluate patients presenting with cognitive concerns, and how that correlates with current research and the established evidence base for the use of such assessments.

Through a survey with general practitioners, these issues are explored:

a) How GPs' use brief cognitive assessments as part of their decision making to identify patients likely to have as-yet undiagnosed dementia;

b) The most commonly-used brief cognitive assessment tools; and

c) How information from the assessment affects GPs' management decisions.

These insights help build a better picture of how GPs choose and use brief cognitive assessments as part of the process for identifying patients with probable dementia in general practice, to better inform policy makers, test developers, researchers and healthcare professionals and help primary care to support a more effective route to dementia diagnosis.

## 6.2. Methods

### 6.2.1. Design

Following a cross-sectional survey design using purposive sampling, participants were presented with a self-completion questionnaire of ten questions and several sub-questions. Questions took three forms: fixed response alternatives; statements where agreement is marked on a five point Likert scale (strongly disagree/disagree/neither/agree/strongly agree); and open free-text comments. Respondents were allocated a number to be used in coding and analysis, as well as to allow differentiation between respondents in any illustrative quotations from free text responses.

The questionnaire was designed for and distributed using Bristol Online Surveys (BOS - www.onlinesurveys.ac.uk). This is a stable and robust academic and research survey tool developed by the University of Bristol. All data is stored securely by JISC hosted in a secure cloud based infrastructure with servers located in the Republic of Ireland, ensuring data is held and processed in compliance with UK Data Protection legislation.

Survey data are encrypted and survey responses were sent over an encrypted SSL connection. BOS user passwords are encrypted using PBKDF2 with a SHA256 hash and a random salt, which fundamentally is a high level of security in web data transfer [344].

### 6.2.2. Participants

All UK general practitioners (including those employed on part time, full time and sessional bases) were eligible to complete the questionnaire. As the population is opportunistically drawn from professional and personal networks within the wider PhD research team as well as via Twitter, the likelihood was always going to be that the survey would identify a greater proportion of people recruited from the South West of England. This is evident in Figure 19 showing clustered locations of general practices

associated with survey participants, but the distribution of respondents is greater than imagined – it is lively that this is mainly due to respondents reached via Twitter.

Participants were informed that by continuing they consent to taking part in the survey, and they could withdraw at any point without giving a reason. A contact email address was given for any concerns or questions participants may have had.

Respondents were informed that all participant data would be anonymised and potential identifiers (such as age, years in practice) were banded into ranges. No names were recorded unless feedback was asked for by respondents (which it was not), and personal data included in these responses would kept confidential beyond the immediate data analyst. In the event no feedback or correspondence was received from individuals, so this was not an issue.

### 6.2.3. Questionnaire development

In consideration of GP time pressures the questionnaire was designed to take less than 6 minutes to complete. The questionnaire was constructed iteratively with an advisory panel of clinical and academic experts, and was piloted with research colleagues to ensure it ran smoothly and functioned as expected.

Initial survey questions were developed through detailed critical reading of the evidence base including key research in the field of brief cognitive assessment use for identifying dementia in general practice[10,15,17-19,21,27,42,43,49,72,92,134,329,331-333,335,336,345-353]. These were developed and refined with close reference to the aims of this project, in order to address the question "how do general practitioners choose and use brief cognitive assessments in practice as part of the process for identifying possible dementia in primary care?". Once questions were structured in a logical way to allow flow from one topic to another and maintain interest across the questionnaire, these were discussed with PhD supervisors and selected Project Advisory Group members consisting of two general practitioners (Willie Hamilton and Nick Cartmell) and one qualitative researcher and research ethicist (Anne-Marie Boylan). Through this process a number of questions were removed and the wording of others was refined in order to improve comprehension and maximise the usefulness of the responses.

The questionnaire was revised further following testing with practicing general practitioners (those named, plus Sam Creavin in Bristol) to identify any GP-specific issues such as difficulties with terminology, jargon or interpretation. All feedback was

used to improve the questionnaire structure and content. Finally, feedback received from the Chair of the University of Exeter Medical School Ethics Committee during the Ethics review process (see Ethics application form in Appendix 14 and Ethics Committee certificate in Appendix 15) was incorporated to further improve aspects on consent and storage of survey data.

### 6.2.4. Procedure

The survey was distributed using the Contact, Help, Advice and Information Network (CHAIN - http://www.chain-network.org.uk) which has approximately 170 general practitioners on the mailing list. Professional networks such as the South West Clinical Research Network and the Exeter Collaboration for Academic Primary Care (APEx) research group helped to publicise the survey. The survey was promoted within the regular newsletter of the South West Clinical Research Network (SWCRN), as well as on Twitter. Retweets were requested from the Royal College of General Practitioners (@RCGP), Devon Local Medical Committee (@Devon_LMC) and practitioner magazines such as GPOnline.com (@GPonlinenews) and PulseToday (@pulsetoday), as well as from individuals associated with those groups and with general practice as a whole.

Originally it was anticipated that the survey would reach around 200 people, based on estimates of reach for Twitter, CHAIN and individual contacts. Once the survey had been launched, this was revised to an estimated 100 people due to the timings of holidays and workload pressures, with an anticipated response rate of around 50%, based on previous experiences of colleagues who have undertaken similar work with comparable target groups [354-356].

## 6.3. Survey analysis

A copy of the questionnaire is available in Appendix 13.

Data were collected anonymously via the Bristol Online Survey tool and analysed using Microsoft Excel 2013 and Microsoft Word 2013.

Quantitative data are reported via ranked preferences of test, with percentage groupings with confidence intervals reported. Univariate descriptive analyses are used to explore demographic variables and GP attitudes to the use of cognitive assessments in general practice.

## 6.4. Free text responses

Within the free text responses in the survey, the coding strategy was purposively developed using emergent themes rather than preconceived ideas[357]. This is not to overlook the place of the author's preconceptions and bias inherent in identifying these emergent themes, and within the discussion section of this chapter there is reflection on the influence of the researcher in surfacing particular elements within the data.

First level coding was carried out manually with pen and paper, and identified a number of common words, phrases and ideas within the GP responses. These codes were refined through revision of the texts within context of first level codes, and discussion of observations with Project Advisory Group members expert in qualitative analysis and survey design. Themes were developed through this process of revisiting and refining the codes, using Excel and Word to create data tables which allowed identification, allocation and grouping of codes from which themes were drawn[358].

After analysing emergent themes at their different levels, a thematic network was developed to structure and characterise these themes. After the network had been created, it was then possible to explore and summarise the thematic network in order to finally interpret patterns discovered within the data[359]. This process drew on expertise and experience within the wider research team to test and refine the analytical framework within which themes are developed. This follows a similar approach to one taken by a member of the Project Advisory Group analysing free text comments from GPs about stroke assessments[360] where data were collated and analysed as previously established for survey responses[361]. In the study on stroke assessment, the analyst [AMB] assigned codes to each comment to explain its meaning and coded extracts were further explored to derive themes from collected responses, using NVivo 10 software to manage data and coding. The reporting of qualitative research follows the recommendations in the Checklist for Reporting Results on Internet E-Surveys (CHERRIES)[362] .

## 6.5. Overview of responses

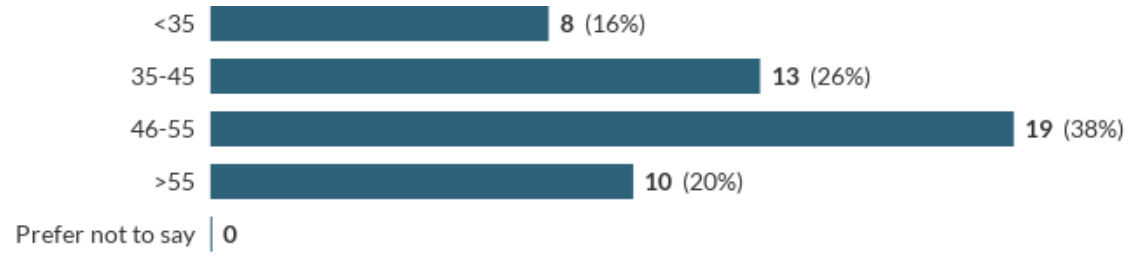
The survey was open from 30th June until 8th September 2017, and distributed via email, Twitter and word of mouth. Fifty-two responses were received in total, and all respondents completing the survey in full. To be eligible, people had to be actively working in UK general practice as general practitioners (GPs – sometimes referred to as family doctors or primary care practitioners).

All survey participants answered all questions fully, and whilst anyone with questions or issues was asked to contact the candidate, there were no reported issues or problems with the survey or the questions posed.

### 6.5.1. Age of respondents

Of the fifty-two people who responded to the survey, the majority (64%) were between 35 and 55 years old. Sixteen per cent of people were less than 35 years old (8/52) and 20 per cent were over 55 years old (10/52). Age ranges are shown in Figure 17.

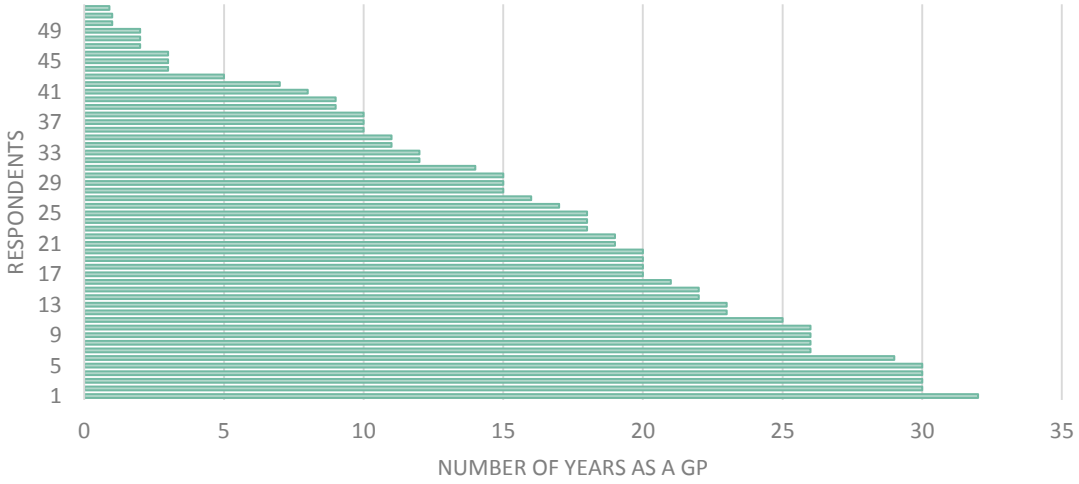
Figure 17 Age of respondents




### 6.5.2. Number of years in general practice

Twenty seven per cent (14/52) of GPs who took part in the survey had been in practise for less than 10 years. Thirty-five per cent (18/52) had between 10 and 20 years' experience, and 39 percent (20/52) had over 20 years' experience in general practice, with one GP having been in practice for 32 years. All responses are shown in Figure 18.

Figure 18 Years since first GP appointment




Geographically, participants were distributed across the United Kingdom with GPs based in general practices across Scotland, Northern Ireland, Wales and England. There was greater representation from GPs based in Devon (31%; 16/52) and in the

South East (23%; 12/52), and lower representation from Wales and the north of Scotland. Clusters of respondents by geographical area are shown in Figure 19.

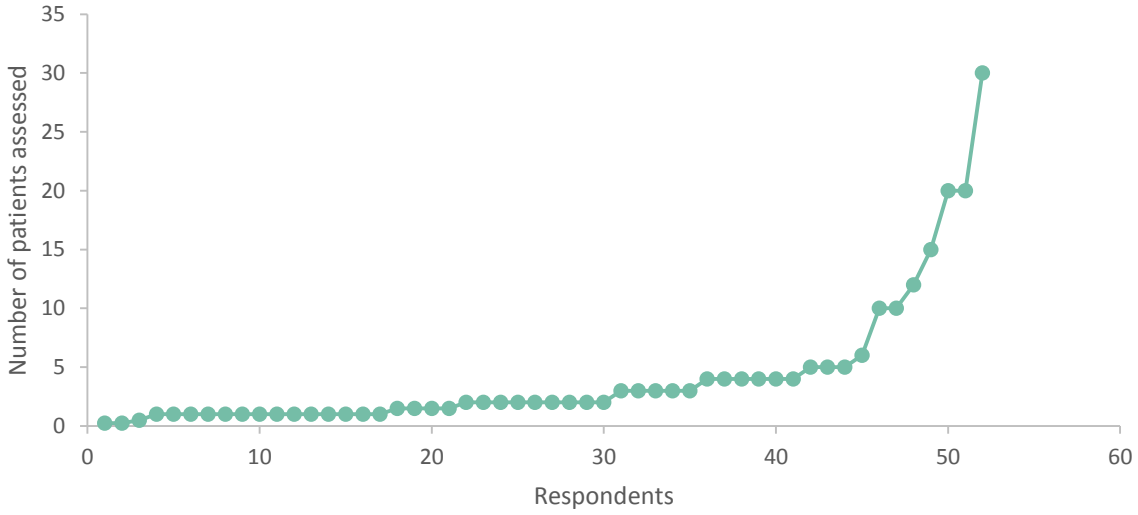
Figure 19 Clustered locations of survey participants by general practice

### 6.5.3. General practice details

Eighty-nine per cent of GPs surveyed did not have access to dementia specialists in their practice. Six GPs reported dementia specialists in their practice, which included a community dementia nurse, a dementia liaison nurse, a dementia service run by a foundation trust co-located within the same building as the general practice, a Primary Care Memory Assessment Service run under a Local Enhanced Service (LES) agreement, and three GPs who are dementia leads or have a special interest in dementia.

The number of patients assessed for cognitive impairment by GPs per month was highly varied, with a mean number of 4 assessments per month (SD = 5.74). Some GPs reported assessing one patient for cognitive impairment every 4 months, whilst others reported assessing 30 patients a month.

Figure 20 Number of patients assessed for cognitive impairment per month




## 6.6. Brief cognitive assessment selection and use

### 6.6.1. The most commonly-used assessment tools

There were two brief cognitive assessments that was most commonly-used amongst the surveyed GPs: the Mini Mental State Examination (MMSE) at 32%, and the General Practitioners assessment of Cognition (GPCOG), also at 32%. The third most-frequently used test was 6-CIT chosen by 17% of respondents. The Clock Drawing Test (CDT - used as a discrete assessment tool rather than as part of another test) was chosen by 5 respondents (7%), and the MiniCog, Montreal Cognitive Assessment, Addenbrooke's Cognitive Assessment third edition (ACE-III) and Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE) were each chosen by 1 respondent (1.4%).

GPs were able to select more than one test, and the most popular combination was MMSE and GPCOG chosen by 8 respondents (15%) with 3 of these GPs adding the CDT, one adding the 6CIT and one using the "Addenbrooke's scale" for younger patients (assumed to be ACE-III). Two of the GPs who selected GPCOG also selected CDT, IQCODE and Bristol Activities in Daily Living (B-ADL). One respondent indicated that they did not have a brief cognitive assessment they used most often.

### 6.6.2. Factors affecting tool selection

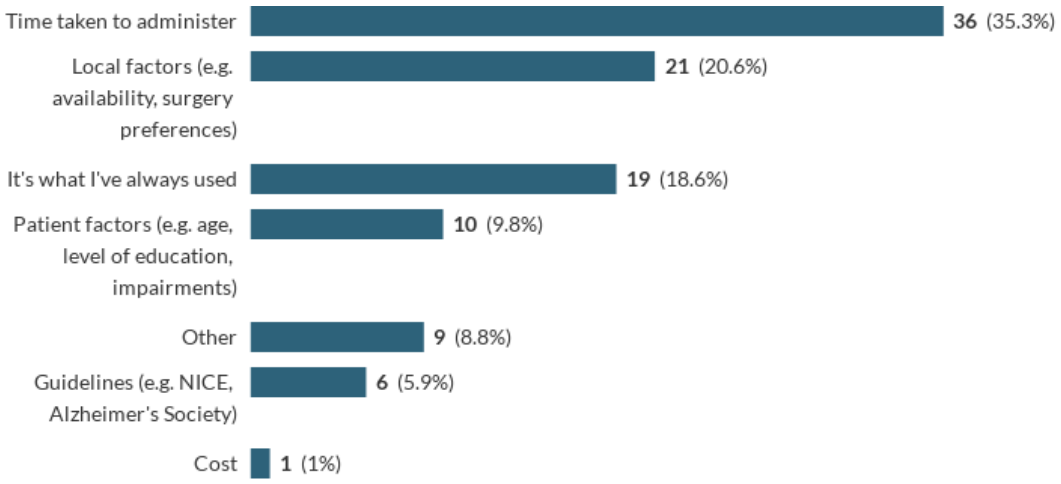
Respondents were asked "what guides your choice of brief cognitive assessment?", with multiple choice options shown in Figure 21. "Time taken to administer" was the most popular answer from 36% of GPs, and "cost" was only selected by one respondent.

Of the 9 GPs who selected other reasons, 3 people had different tools embedded in their clinical EMS software (used within a number of practices of respondents), and one GP had taken advice from a specialist colleague. One GP noted that their tool of choice (GPCOG) was needed for memory clinical referral. One GP commented that they felt the ACE-III has good face validity and they felt a normal range score on ACE-III offered greater reassurance:

> *"… [A] normal-range score is more reassuring (as important as its use in identifying actual cognitive impairment)" GP respondent-1*

One GP noted the value of a test (GPCOG) in allowing the involvement of a relative if needed, and one participant mentioned possible barriers to tool selection: remembering its' name; and finding it on the system.

Figure 21 Factors affecting GPs' choice of brief cognitive assessment



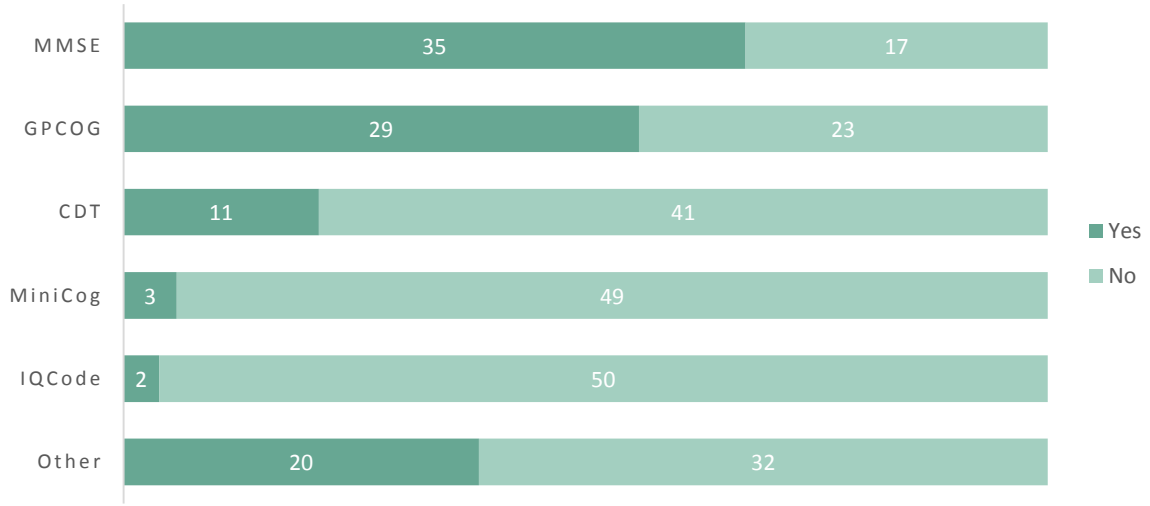

## 6.7. Brief cognitive assessment use as part of GP decision making to identify dementia

### 6.7.1. Use of brief cognitive assessments in practice: for initial assessment; and for monitoring over time

GPs were asked which of the tools they had used *in the last two years* for a) initial assessment of a patient (see Figure 22); and b) decline over time.

Tests were most often used for initial assessment of a patient, and tests were less often used for monitoring decline in cognition over time. The majority of respondents (67%) used MMSE for initial assessment, and MMSE was also the most popular for monitoring decline over time (40%).

Figure 22. Have you ever used this brief cognitive assessment for initial assessment?

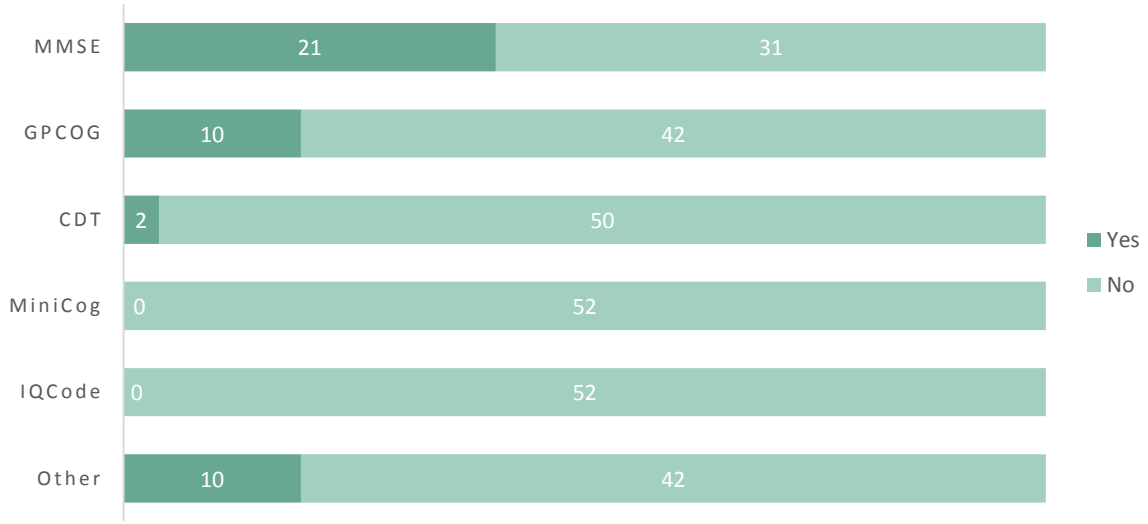



Within the 'Other' category, 13 out of 52 GPs (25%) had used 6CIT for initial assessment of patients, with three others using ACE-II (6%), two GPs (4%) referring to the Montreal Cognitive Assessment (MoCA) and Bristol Activities of Daily Living and AMT each referred to by one GP (2%).

Fewer GPs had used brief cognitive assessment tools for monitoring decline over time (see Figure 23) but the same pattern remained as in conducting initial assessments, with MMSE most popular (40%), followed by GPCOG (19%), CDT (4%) and 6CIT (under 'Other' – 14%). Also under 'Other' two GPs selected ACE-III (4%) and one GP selected MoCA (2%) for monitoring decline over time.

IQCODE and MiniCog were not selected by any respondents for monitoring decline over time.

Figure 23. Have you ever used this brief cognitive assessment for monitoring decline over time?




### 6.7.2. Acceptable test administration time

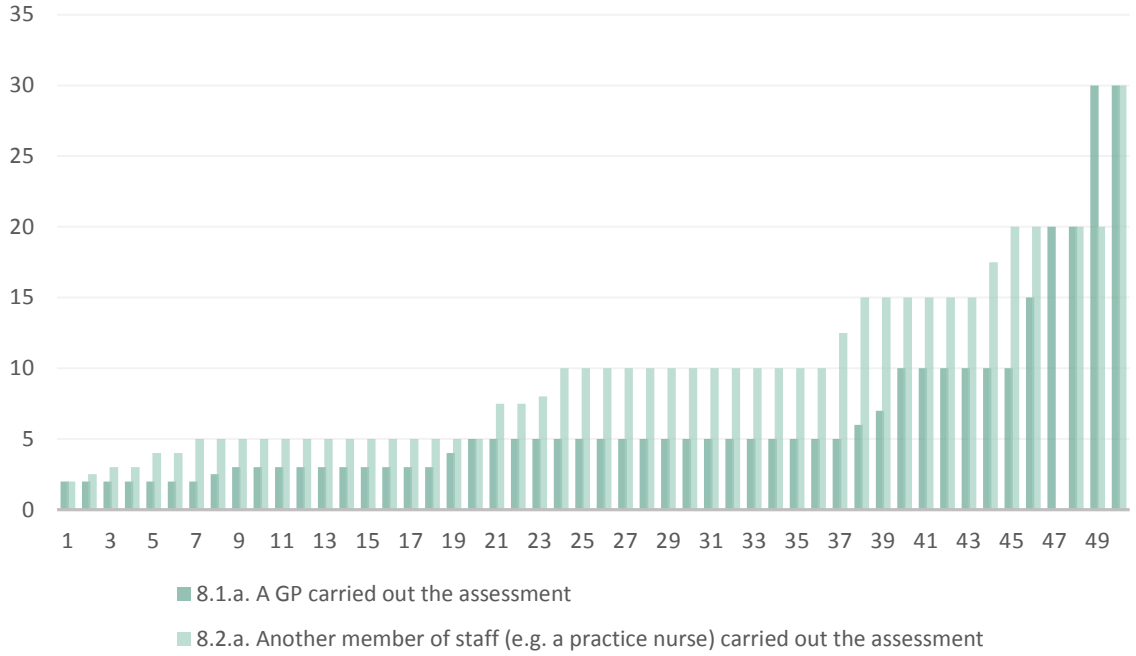
GPs were asked what they would consider an acceptable administration time of a brief cognitive assessment within general practice in two different scenarios: if they themselves carried out the assessment; and if another member of staff such as a practice nurse carried out the assessment. The data were entered into a free text box, so GPs were free to choose any number or add other figures if they chose to.

Of the 50 eligible GP responses regarding test time if a GP carried out the assessment, 14% stated an acceptable test time of 2 minutes, 22% stated between 2 and 3 minutes, and 38% chose between 4 and 5 minutes (one GP chose 4 minutes and the rest chose 5 minutes). One GP selected 6 minutes, one chose 7 minutes, six (12%) selected 10 minutes, one chose 15 minutes, two (4%) selected 20 minutes and two (4%) selected 30 minutes. The range of scores are shown in Figure 24. Out of the 52 respondents, one GP entered ineligible data in this section ("yes" rather than a number) so their responses could not be counted. One GP entered a question mark ("?") with the comment "Not really sure. I suppose less than 10 mins ideally".

The accompanying question "how long would be an acceptable test time if another member of staff (such as a practice nurse) carried out the assessment?" had 49 eligible responses. In addition to the two ineligible replies above which were repeated across both questions, one other respondent had typed "practice nurse" within this section rather than giving a number. Of the 49 eligible responses, several GPs had entered a range rather than a single figure (e.g. 10-20 minutes). For analysis, the mid-

point of the range was taken (e.g. from 10-20 minutes this was 15). Combined data are shown in Figure 24.

Figure 24 Acceptable brief cognitive assessment administration time if: (see legend)




It is clear from Figure 24 that on the whole, GPs felt it acceptable for another member of staff such as a practice nurse to spend longer administering a brief cognitive assessment than a GP.

Comments made reinforced this finding, and indicated other areas of influence:

*"Needs to be consistent within the practice, not have different tests used by different staff" [GP respondent-3]*

*"if shown to be more reliable test" [GP respondent-46]*

Many participants highlighted common subjects of time, efficiency and system factors:

*"[another member of staff would be] able to have longer as can concentrate on assessment only" [GP respondent-17]*

*"If booked as a separate appointment" [GP respondent-26]*

*"GP's time probably more productively spent doing something else and anyone trained could do [it]" [GP respondent-29]*

Funding or financial issues were also consistently mentioned:

*"again, the nurse would usually take more time over most things but we pay for nurse appointments and time and they are part of the*

*total pressure on the system, so nurse appointments are not resourced" [GP respondent-28]*

## 6.8. Confidence in practice

### 6.8.1. GP confidence in interpreting test results

GPs were asked how confident they would be in interpreting the results of a specific named test, on a scale of 0 to 5 where 1 was 'not at all confident' and 5 was 'very confident'. For GPs who selected MMSE as the brief cognitive test they used the most (N=22), results of confidence in interpreting results are shown in Figure 25 with the legend showing the level of confidence (2-5). For those GPs who selected GPCOG as the brief cognitive test they used the most (N=18), results of confidence in interpreting results are shown in Figure 26 with the legend showing the level of confidence (2-5).

Figure 25. GP confidence in interpreting the results of MMSE [N=22]

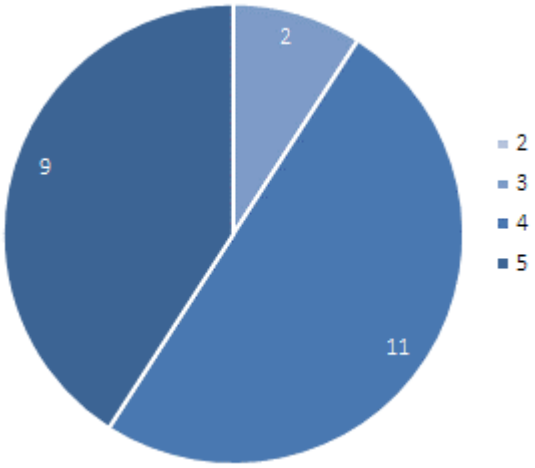
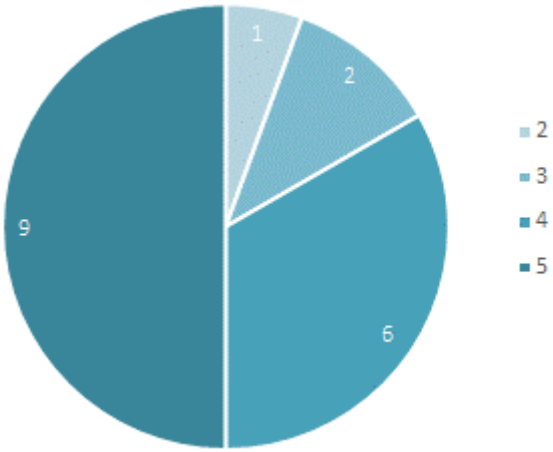



Figure 26. GP confidence in interpreting the results of GPCOG [N=18]




Nobody responded with 1 (not at all confident); one GP rated their confidence in interpreting GPCOG as a 2 (not very confident), and added the following comment:

*"it [MMSE] is such a basic scale it can be skewed by anxiety/a bad day etc., and it doesn't pick up subtle issues" [GP respondent-28]*

Other participants rated their confidence in interpreting results of the MMSE at 4 (fairly confident) or 5 (very confident), and there were comments on the difficulty of interpreting results that were closer to the normal thresholds or bias may be an issue:

*Language can be issue. Very intelligent folk score well even when in decline so [I] feel less confident then [GP respondent-20]*

### 6.8.2. GP confidence in explaining test results to patients

GPs were asked how confident they would be in explaining results of a specific (named) test to patients using the same scale described above, and results are shown in Figure 28 for those GPs who selected MMSE [N=22] and Figure 29 for those GPs who selected GPCOG as their most frequently used test [N=18].

Figure 28. GP confidence in explaining the results of MMSE to patients [N=22]

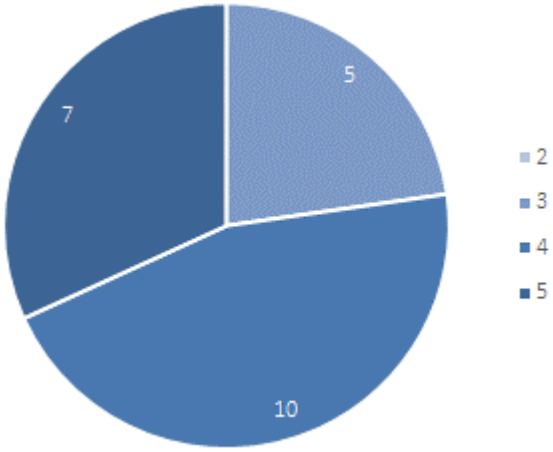
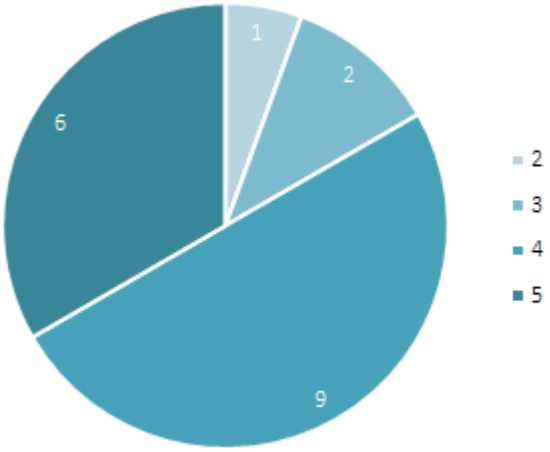



Figure 29. GP confidence in explaining the results of GPCOG to patients [N=18]




Nobody responded with 1 (not at all confident), but one GP rated their confidence in explaining the results of the GPCOG to patients as a 2 (not very confident). This respondent was the same person as had responded with a 2 to the previous question asking how confident they were in interpreting results. This respondent made the following comment:

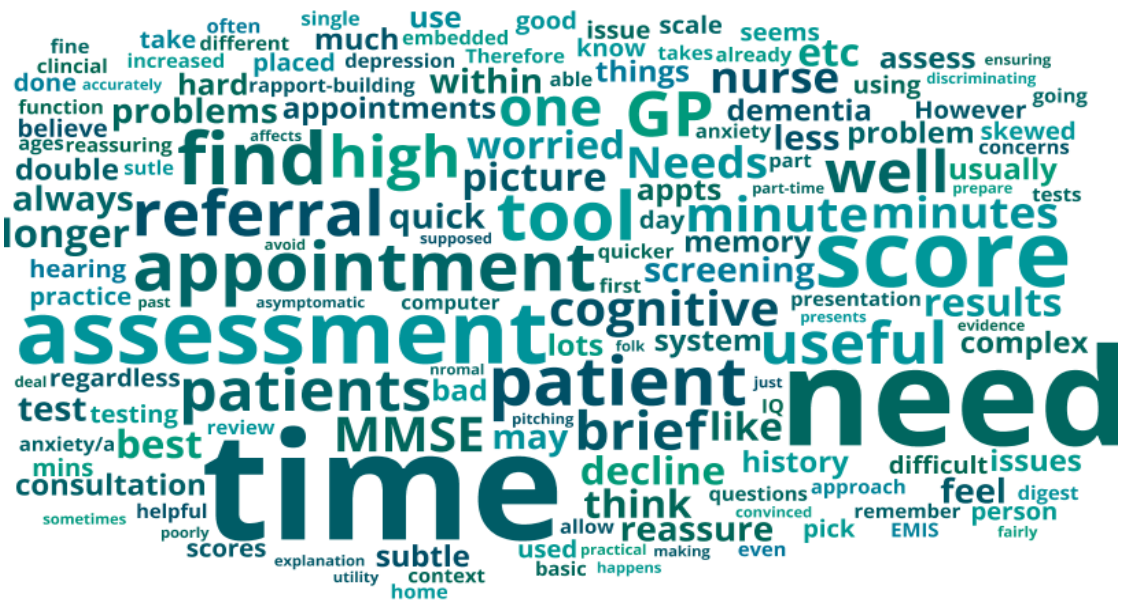
> *"Able to reassure if normal (and if clinical picture is one of worried well) but not if more complex picture /subtle problem or high IQ. Also useful if low score to say - needs referral for further testing - looks like a possible memory issue."*
>
> *[GP respondent-28]*

More GPs responded to this question with a 3 (moderately confident) but did not explain this result further. The final question (Question 10) within the survey asked GPs for any further comments about brief cognitive assessments, particularly in relation to identifying possible dementia. Twenty out of the 52 respondents chose to comment at question 10.

## 6.9. Analysis of free test responses within the questionnaire

As a first step in assessing the responses within all the free text sections in the questionnaire, a word cloud was produced to visually explore the content of responses for frequency of terms, concepts and words used by respondents. These word clouds are in essence a list of words weighted by frequency through a simple algorithm and used primarily to visualize the frequency distribution of metadata entered – in this case, the free text responses given within this questionnaire. It is designed as a visual check to detect potential patterns and associations between different words. As shown in Figure 30, 'Time' and 'need' were the most frequent terms used, with assessment-related terms (such as 'tool', 'score', and 'assessment') also appearing frequently.

Figure 30 Word cloud of all free text participants responses




Initial coding of data was carried out by creating segments of text fragments focussing specifically on the aims of the survey: to investigate current clinical practice in the use of brief cognitive assessments to evaluate patients presenting with cognitive concerns, and how that correlates with current research and the established evidence base for the use of such assessments.

These codes were then critically assessed and re-read in the context of the survey text segments, extracting and refining emergent themes as they developed[359]. Following discussion of these developing themes within the research team, themes were arranged and summarised to reflect the principal findings within the data. A simplified table showing the process of coding and theme development is available in Appendix 16.

These themes are discussed in detail below.

## 6.10. Barriers and facilitators for practice identified by respondents

Responses reflected a view of different factors as barriers or facilitators to clinical practice. These elements could be assessed in a number of ways, but the themes split simply into GP factors, patient factors, test factors and system factors.

### 6.10.1 GP factors

A number of GP responses suggested confidence in their clinical decision making which would influence how they conduct and interpret the results of an assessment. Some GPs voiced scepticism in the ability of brief cognitive assessments to influence their clinical judgment:

*"Most patients score fairly poorly on it and need referral regardless."*

*"…if I have enough clinical suspicion, I will refer regardless of high score" [GP respondent-10]*

Some respondents reflected on the importance of interpersonal dynamics between patient and practitioner in conducting the assessment:

*"People can be really upset (angered) by the dementia screening questions. It's worth not rushing." [GP respondent-43]*

*I find that patients on first presentation may need rapport-building. Often hard of hearing. Need time to digest what's going on. However, what usually happens is that there is almost always a social or physical need that presents at the same time. They present because of a problem." [GP respondent-10]*

One respondent questioned the role of the GP, although it was unclear whether this was around the diagnosis of dementia, use of brief cognitive assessments or another element:

*"As a GP I don't believe that I am the best placed person to do it." [GP respondent-47]*

### 6.10.2 Patient factors

A number of GPs highlighted their prior knowledge of the patient as having an impact on how the brief cognitive assessment was used. Although it was not clear how this prior knowledge would influence them, this familiarity was sometimes paired with discussion of cognitive decline:

*"Takes into account prior knowledge of the patient" [GP respondent-39]*

*"It depends if you know the patient already & have identified cognitive decline" [GP respondent-34]*

Patient behaviour and particularly the tendency to present to general practice with several issues was remarked on as a factor in how GPs carried out brief cognitive assessments:

*"We are often trying to implement these with an aging population, who need time, sufficient hearing and motor function to complete many of them. They often have many issues in the consultation" [GP respondent-10]*

Finally, the perceived role of the GP in reassuring patients was commonly mentioned:

*"Able to reassure if normal (and if clinical picture is one of worried well) but not if more complex picture /subtle problem or high IQ" [GP respondent-28]*

*"I do the longer Addenbrooke's one for high IQ/young patients but it takes ages and feels like a hassle but can be useful to reassure worried well and avoid a referral" [GP respondent-29]*

### 6.10.3 Test factors

Common characterisations assigned to tools were either assistive, using language such as "helpful" or "useful", or flawed, with language such as "skewed" or "crude":

*"It is such a bad scale it can be skewed by anxiety/ bad day etc. and it doesn't pick up subtle issues"[GP respondent-28]*

*"Clock drawing often really helpful" [GP respondent-29]*

*"…the quicker ones are quicker but less useful" [GP respondent-28]*

The length of time that a test takes was a popular and recurring theme across the survey. The need for a test to be short, quick or to fit within a standard consultation time was the point most frequently made by GPs:

*"If you only have 10 mins, the assessment is the thing that gets squeezed" [GP respondent-10]*

*"I think a brief assessment tool should be just that, brief" [GP respondent-17]*

*"Needs to be quick" [GP respondent-21]*

### 6.10.4 System factors

Several respondents mentioned the computer systems within their general practice – either EMIS or SystemOne. A number of these systems have one or more brief cognitive assessments incorporated as part of the general practice software, which allows GPs to use specific tools such as 6CIT directly from the computer with scoring and assessment conducted automatically.

*"I have personalised a version of MMSE and embedded it within EMIS… This makes it easy to administer." [GP respondent-39]*

*"Available on EMIS system" [GP respondent-31]*

*"It's written on the screen... would I remember scoring? Not a chance but I don't need to." [GP respondent-43]*

Other responses highlighted the influence of test availability and increased challenges following enforced copyright of the MMSE:

*"Stopped using MMSE when charges introduced for using it" [GP respondent-3]*

*"Can be very time consuming. Also MMSE now copyrighted..." [GP respondent-18]*

Time pressures were also commonly mentioned, with some GPs commenting on how considerations of time allowed within the general practice consultation affected their clinical decision making:

*"Pressure on time and appointments drives lots of GP decision making" [GP respondent-28]*

*"They need to be short in time to facilitate a realistic assessment in general practice" [GP respondent-1]*

*"Lots more to do in the 10 minute consultation" [GP respondent-46]*

Other members of the general practice such as nurse practitioners were viewed as having more time to conduct brief cognitive assessments and – by one GP respondent – was viewed as conducting brief cognitive assessments more accurately:

*"Done more slowly and accurately by the nurse" [GP respondent-52]*

*"Again nurse would usually take more time over most things…" [GP respondent-28]*

However these members of staff were also viewed as under-resourced, and possibly less skilled:

*"Nurse appointments are not resourced" [GP respondent-28]*

*"GP's time probably more productively spent doing something else and anyone trained could do" [GP respondent-29]*

Some GPs also emphasised the importance of considering the whole process of assessment beyond brief cognitive testing:

*"The test is usually for someone else's need/purpose. Diagnosis is a shared interpretive process of weighting up risk and benefit with patient" [GP respondent-48]*

*Value judgments*

The language used around brief cognitive assessments and the testing process was often weighted; tests were good or poor, helpful and useful or flawed and crude. Respondents used language with comparative undertones of quality or worth:

*"Need the best tool available" [GP respondent-5]*

*"No perfect one!" [GP respondent-42]*

*Quality versus efficiency*

One GP's comments carried an implicit suggestion that brief cognitive assessments might replace or overrule other aspects of general practice assessment such as history taking:

*"They can be useful, but as with depression scores, they're no substitute for a good history. The danger can be that they lead to a tick box approach" [GP respondent-2]*

This caution against taking a 'tick box approach' contrasts with comments on balancing the demands for quality – and particularly accuracy - in the brief cognitive assessment with efficiency and speed of process.

*"Not always accurate" [GP respondent-21]*

*"Needs to be fitted into surgery time" [GP respondent-18]*

*"Needs to be 'brief'" [GP respondent-1]*

*Generalism versus specialism*

The place of specialist skills within general practice was another theme that emerged amongst the responses; one GP referred to getting "advice from a specialist colleague" [GP respondent-27] and lack of expertise was referenced by another couple of respondents:

> *"Done more slowly and accurately by the nurse" [GP respondent-52]*

> *"I have not seen much cognitive decline. Working part-time may contribute to this." [GP respondent-1]*

The role of GP as gatekeeper or enabler for patients to be referred on for further assessment was referenced obliquely by other respondents:

> *Able to reassure if normal (and if clinical picture is one of worried well but not if more complex picture /subtle problem or high IQ" [GP respondent-28]*

> *"…it's a case of "this seems fine we need to think more about your problems (or not)" or "this suggests you would benefit from seeing someone else to review things further." [GP respondent-43]*

One GP clearly queried the role of the GP in cognitive assessment:

> *"As a GP I don't believe that I am the best placed person to do it [cognitive assessment]." [GP respondent-47]*

Other GPs questioned the place of screening in general practice:

> *"I'm not convinced of the evidence base or practical utility of asymptomatic screening" [GP respondent-5]*

> *"I don't like the screening aspect. I prefer to use [the test] when someone is worried about their memory" [GP respondent-7]*

*Evidence-based practice*

The question which this survey sought to address is: how do general practitioners choose and use brief cognitive assessments in practice as part of the process for identifying possible dementia in primary care? There are clearly many systemic factors which may contribute to constraining GP practice and are referred to by GPs responding to this survey. These potential constraints include length of time of consultations, funding for other staff such as nurse practitioners, and the use of practice-based computer systems with in-built software.

Areas which were less well-characterised by respondents than had been anticipated by reading the wider literature were levels of experience related to patterns of working (full time versus part time, partnership versus locum status), low reported numbers of dementia speciality within general practices, and sparse mention of relatives or carers – particularly given the popularity of the GPCOG, which is often preferred as it contains a section for assessment via an informant.

## 6.11. Summary of main findings

The MMSE and the GPCOG were the most frequently-used tests according to respondents, and were equally popular with 32% of those surveyed using one or the other (or both) most regularly. CDT, 6CIT, ACE-II, IQCODE and the Bristol Activities in Daily Living tool were also mentioned.

Of the factors which guided tool selection, 'time taken to administer' was the most prominent reason to choose a test (chosen by 35.3%), followed by 'local factors' (chosen by 20.6%), including availability of the test and preferences of the individual surgery. Nineteen respondents (18.6%) selected 'it's what I've always used', with 9.8% of people choosing 'patient factors' such as age, education or impairments.

The majority (67%) of people surveyed used MMSE for initial assessment of a patient, and 40% of people had used MMSE to monitor decline over time. Fewer respondents generally had used any of the tests to monitor decline over time, and neither IQCODE nor MiniCog were used by any contributors for monitoring decline over time.

Thirty-eight per cent of respondents felt that the optimum length of time for a test to administer was between 4 and 5 minutes. Those who responded consistently felt that another member of staff (such as a practice nurse) could reasonably spend longer administering a brief cognitive test compared to a GP.

When asked how confident they were in interpreting results on a scale of 1-5, with 5 being most confident, 11 out of 22 (50%) who used the MMSE felt they were fairly confident (4 out 5) and 9 out of 22 (41%) were very confident (5 out of 5). For GPCOG, of the 18 respondents, 9 (50%) were very confident (5 out of 5) in interpreting results, 6 (33%) were fairly confident (4 out of 5), two (11%) were quite confident (3 out of 5) and one was not very confident (2 out of 5).

In terms of confidence in explaining test results to patients, patterns were similar with 10 out of the 22 people who used the MMSE (45%) being very confident in explaining

results to patients, 7 out of 22 (32%) fairly confident and 5 out of 22 (23%) were quite confident. For the 18 people who chose the GPCOG, 6 out of 18 (33%) were very confident in explaining results to their patients, 9 out of 18 (50%) were fairly confident (4 out of 5), two (11%) were quite confident (3 out of 5) and one was not very confident (2 out of 5).

Within free text responses, GPs noted scepticism in the ability of brief cognitive assessments to influence clinical judgment, identifying the doctor-patient relationship as more important and voicing doubts over whether the GP was the best person to conduct a brief cognitive assessment. The role of the GP in ressuring the patient regarding test results and managing complex conditions was a common theme. Tests were characterised as assistive or flawed, and the length of time a test takes to administrate was a recurring topic. Systemic factors were seen as helpful (with some tests embedded within the computer system) or obstructive (such as time pressures, costs for using tests and under-resourcing of other practice staff). GPs questioned the role of screening in general practice, and several respondents referred to tensions between specialist skills and the role of the generalist in general practice.

## 6.12. Threats to validity

Most responses were from England, but replies were received from Northern Ireland, the Scottish Highlands and North Wales. Geographically respondents were fairly well dispersed, with clusters (suspected to be around prominent respondents who would have forwarded on requests to colleagues) around London, Bristol, Birmingham, Exeter and Torquay. Other demographic details such as sex and ethnicity were not collected, as these were not *a priori* motivations of the survey and these data were not guaranteed to be used within the final analysis.

Data were not collected on career breaks or whether hours were full or part time. Whilst this may have been interesting in assessing how work patterns related to the reported number of patients assessed for possible dementia per month, the survey was designed to be quick and easy to complete with as small a number of questions posed as would be informative to the research. The central research aims were to discover: how GPs' use brief cognitive assessments as part of their decision making to identify patients likely to have as-yet undiagnosed dementia; the most commonly-used brief cognitive assessment tools; and how information from the assessment affects GPs'

management decisions. Therefore any non-essential items (i.e. not relating to the core research aims) were removed during the refining and development process.

By employing an opportunity sample publicised using professional networks, social media and word of mouth, a fair response rate of 52% (52 out of an expected 100 people) provided a rich source of data for analysis. This sampling framework does introduce a number of potential biases. All participants are self-selected, volunteered their time and are active on Twitter. It is most likely that respondents are actively interested and engaged in dementia diagnosis within general practice, as they agreed to take part in a survey on this topic for free. Many respondents were colleagues and referrals on from people either within the PhD project team or GPs who had piloted the survey. The topic itself and the nature of self-report may have introduced social desirability bias, where respondents were motivated to appear engaged, active and 'doing the right thing' with their own practice regarding dementia diagnosis. Questions were designed to capture a range of responses and avoid explicit bias towards one course of action, but this element of desirability may not have been eliminated.

## 6.13. Triangulation of the data, incorporating perspectives from other parts of the thesis and existing literature

In order to assess the richness and authenticity of these findings, and – where possible – to generate innovation in framing of concepts, cross verification of the survey data was conducted with evidence from the rapid review, overview and systematic reviews conducted as part of this thesis, as well as with the known body of existing literature around GP selection and use of BCAs when used as part of the process for identifying dementia in general practice[363].

This triangulation of the survey data provided a rare and valuable opportunity to provide new insights and deepen understanding of the existing evidence base alongside new information generated through the GP survey[364].

Triangulation was achieved through assessment of these survey findings alongside results of the rapid review of clinical practice guidelines, the overview of BCAs available for use in general practice, the systematic review of direct comparisons of the diagnostic accuracy of MMSE and GPCOG when used for identifying dementia in general practice, and a thorough review of the literature around BCA selection and use when applied in a general practice setting for the purpose of identifying dementia as part of the diagnostic process.

Within the rapid review of CPGs, four guidelines were identified that fitted the criteria of identifying themselves as being CPGs, explicitly suitable for general practice, directly focussed to identifying dementia and available in English language.

All CPGs were single national guidelines, published in the last 13 years. One was from the UK[173], one from Australia[174] and two from Canada[110,175]. All included CPGs were created for an audience wider than general practice, including hospitals, community care, specialist assessment, and care homes. There were a number of guidelines within all three CPGs not applicable to general practice.

In terms of procedure for conducting diagnostic assessments, the UK CPG recommends that at the initial consultation a patient history is taken from the individual and *ideally* from someone who knows the person well. If dementia is still suspected the patient should have a physical examination, alongside appropriate blood and urine tests to rule out other causes of cognitive problems, and cognitive tests should be conducted. The Australian CPG are based closely on the UK CPG, but the guideline committee noted caution and adaptation when assessing people from indigenous communities as there may be different perceptions of alterations in cognition in some cultures, leading to individuals and their carers seeking diagnosis at a later stage of the dementia progression. The Canadian CPG does not give guidance on diagnostic procedure in general practice.

These findings were not confirmed or refuted within the survey of GPs, as none of the GPs surveyed mentioned specific CPGs. As the survey was conducted in the UK, only the NICE guidance would be relevant or easily accessible but it is notable that these – or any – named clinical practice guidelines are not specifically referred to in any of the closed response questions or any of the free text responses.

A 2002 survey of clinicians in six United States Department of Veterans Affairs medical centres[365] found that whilst clinician practice broadly followed dementia clinical guideline recommendations, roughly one third of those surveyed (N=200) did not routinely use a recommended BCA. One third of respondents did not routinely discuss care or management needs arising from the assessment with the patient or carers, and one third of clinicians did not discuss the diagnosis or prognosis directly with their patient.

A randomised controlled trial of 35 UK GP practices in 2009[366] found that only 28% of patients with a diagnosis of dementia received BCA testing in primary care, with the test results recorded correctly in only 23% of cases. Of those patients that received a diagnosis of dementia, only 15% of patients were told they had dementia within primary care. This study highlights the potential anomaly that whilst GP records may not have recorded a diagnostic or management process, it may have taken place but not been recorded within standard record-keeping software and other work clearly reinforces this common discrepancy between GP practice and record keeping[17,367-369].

A 2015 analysis of GP records relating to dementia diagnosis in the UK[370] reported general improvements in GP concordance with guidelines on diagnosis and management of people with dementia, following publication of the 2006 NICE clinical guideline 42[371] and the 2014 UK Dementia Strategy[372]. Only 40% of practices surveyed recorded carrying out BCA tests. Of the patients with cognitive tests recorded, 57% were for the MMSE, 36% were for the AMT and 4% were for 6-CIT. This is out of step with the findings of the survey reported here, where MMSE made up 32% of the BCAs reported, GPCOG featured in 32% of responses and 6-CIT in 17% of responses. The Abbreviated Mental Test was not mentioned by any of the GP respondents in the survey conducted within this thesis.

In terms of the actual BCAs identified by GP respondents to the survey, these were broadly in line with those identified in the overview of BCAs available for use as part of the process for identifying possible dementia within general practice. In the overview, MMSE was the BCA which featured most frequently within the included systematic reviews, as it appeared in 8 out of the 13 systematic reviews. GPCOG featured in 5 out of the 13 reviews, with both CDT and MiniCog featuring in 4 reviews each. Within the survey, MMSE and GPCOG both featured in 32% of responses, with 6-CIT appearing in 17% of responses and others (CDT, MiniCog, MoCA, ADAS-Cog and IQCODE) all mentioned by at least one respondent. Within the overview, both ADAS-Cog and MoCA were not included as tests suitable for general practice as they both take longer to administer[42,104] than the 10 minute threshold set within the overview selection criteria as suitable for use in general practice.

One of the main findings of the overview in relation to suitability for general practice was that the MMSE is not suitable as the administration time is consistently longer than the average time allowed for most GP consultations[133,134]. The survey identified

that other BCAs may be growing in popularity compared to the MMSE, and this may be a pragmatic response – at least in part - to increasing demands on the GP consultation and limited time allowed for a number of activities[373], alongside growing GP understanding of bias within the MMSE relating to education and culture[254,264,374].

Within the systematic review directly comparing the diagnostic accuracy of GPCOG and MMSE when used as part of the process for identifying dementia in general practice, it was shown that MMSE and GPCOG Total performed similarly in terms of sensitivity and specificity. This finding was reflected within the survey, where equal numbers of GPs surveyed most commonly chose MMSE and GPCOG for use in their practice as part of the process for assessing patients for possible dementia.

Whilst a major finding within the systematic review was that the threshold selected may have more influence on test accuracy than the test itself, the issue of appropriate thresholds and tailoring of the test did not feature within the survey of GPs. Whilst there was not one question which addressed this explicitly, it may have been expected to have featured within one of the free text responses. It may be that GPs surveyed were not sufficiently familiar with the BCAs to acknowledge the potential variation at different thresholds, or it may simply have been missed amongst other considerations around BCA selection and use.

One insight provided by this survey which was strongly resonant with the systematic review findings is the place and importance of factors beyond classic diagnostic accuracy measures to consider within general practice concerning BCAs. Concerns around pragmatic factors were raised, such as the time allowed within the consultation for actual conduct of the BCA, including explanation to the patient, gathering of pre-assessment information, conduct of the assessment, analysis, explaining to the patient what the results mean and then discussing with the patient and family as appropriate the potential prognosis and options for ongoing treatment and management. Free text responses within the survey identified GP concerns around patient factors including implications of a diagnosis, lack of caregiver support, and not wanting to upset the patient unduly through the process of testing.

These findings resonate strongly with the literature around patient factors beyond accuracy that concern GPs when conducting an assessment for dementia[10,43,69,348]. It has been proposed that in the current model of dementia diagnosis in general practice,

patient preference may be viewed as subsidiary to clinical ability, where the dominant medicolegal framework with government targets and institutional narrative driving early diagnosis discounts the role of the patient in managing their own health journey and right to challenge medical intrusion[375]. Topics related to the patient's ability to function on a daily basis, and the implications of a dementia diagnosis on a patient's daily activities, did not feature within the free text sections of the survey. It was surprising to find how little these wider implications of the diagnosis and understanding of the testing were mentioned by GPs, yet these factors are identified regularly in the literature[16,370,376,377].

There was also scant mention of relatives or carers within the survey responses. This was another surprising finding as the GPCOG was so popular amongst respondents, and is often preferred as it contains a section for informant assessment unlike the MMSE. Research literature makes frequent reference to caregiver perspectives, in terms of perceived support[376], a lack of coordination for care[347], and poor communication within general practice[378-380]. This oversight may have been due - in part at least - to a lack of explicit questioning within the survey directed towards interactions with family and caregivers.

One of the singularly most surprising findings from the overview was echoed within the survey – that there was no evidence for one BCA perceived as most suitable for GP use as part of the process for identifying dementia. The lack of clear evidence at the level of primary studies, systematic review or within clinical practice guidelines reflected strongly in the disunity across GP preferences and concerns around current BCAs used in clinical practice. In addition, the mixed levels of knowledge, understanding and confidence in testing generally and in the choice and application of specific BCAs which was revealed by the survey of GPs can be understood as a direct and logical product of the findings of the overview – that current evidence is unclear, of mixed quality, and not fit for purpose in informing general practice. Equally the rapid review of CPGs relating specifically to the use of BCAs in general practice similarly found a lack of evidence to help guide general practice in choosing and using BCAs appropriate to their individual needs and populations.

In terms of practical improvements, new insights were gained from the survey around the convenience of incorporating assessments within the general practice software, generally EMIS or SystemOne. Respondents indicated the convenience of having

assessments already installed on the system which meant the GP no longer had to find the right assessment or think about what BCA to use. Conversely, respondents to this survey spoke of guarding against a tick-box approach to cognitive assessment. There is little research literature on this topic relating specifically to dementia diagnosis in primary care, but in the area of blood testing a qualitative assessment of patient and clinician perspectives found that there may be value in adopting online systems to reduce areas for error (such as interpretation or transposition) and increase fluency in communicating results to patients and between different clinical areas such as pharmacies and general practices[381]. This may be an area where simple pragmatic changes can significantly improve current practice, but there is work to be done on assessing costs as well as benefits that greater efficiency and convenience may introduce.

This process of triangulation of the survey data with findings from the rapid review, overview and systematic review, alongside investigation of current research, has allowed a deeper and more nuanced assessment of the discoveries made within this thesis. Moreover, the process of triangulation has highlighted areas of resonance as well as contrasts within what is already known and what has been discovered in the course of this thesis, such as the potential for incorporating approved BCAs into GP systems to improve adoption and consistency in recording of assessments. This has enabled further insights to be made, and these are discussed in more detail within the final chapter.

## 6.14. Recommendations to practice

From the results of this survey, there appears to be a limited pool from which GPs are selecting brief cognitive assessments. The continued popularity of the MMSE despite its copyright enforcement now carrying a cost (or risk of legal action) for individual practices is a major concern, both in terms of a lack of evidence-based practice and from a cost-effectiveness perspective. There appears to be a clear deficit in the range of brief cognitive assessments that GPs are considering for practice, and the reasons given for selecting particular tests need further examination. Administration time was the most prominent influence in this decision-making yet the MMSE takes longer than the average consultation length to administer from start to finish[382]. As has been recommended elsewhere in this thesis, establishing actual administration times for the most popular tests is long overdue.

Policy recommendations for conducting brief cognitive assessments in practice must take account of limited time available within the GP consultation, limited expertise of dementia diagnosis and management within practices, and finite resources from other practice staff who may be asked to carry out assessments. In addition, guideline developers and policy makers must recognise the limits of current knowledge in identifying dementia. The judicious use of cognitive assessment should be encouraged alongside other measures particularly for people who may be disadvantaged by common tests, such as those with higher education levels, where English is not their first language, or from diverse ethnic backgrounds. These factors are well recognised within general practice, yet both the tests available and the policies behind current practice often do not take full account of these features.

It was interesting to note that tests that are integrated into the practice software are seen as useful, and this type of service integration to ease administration for GPs and the wider practice should be investigated as a simple benefit for administration and management.

Finally, the increased drive towards asymptomatic screening within the general practice population (‘case finding’, or ‘targeted screening’) is clearly a concern and some respondents voiced discomfort with pressures in this direction. Until the evidence base is better built and understood, policies designed to increase case finding should be suspended.

## 6.15. Recommendations to research

Whilst the popularity of MMSE as a brief cognitive assessment for use in practice is waning following the imposition of copyright regulations meaning tests now have a licensing charge, it is still a highly popular measure within research. This should change, as a matter of priority – particularly for research focussed towards primary care and general practice.

Whilst GPCOG was also popular amongst survey respondents, these two tests may not suit the general practice setting where short consultation times mean tests need to be shorter than both of the most popular measures within this survey (MMSE and GPCOG). The development and evaluation of brief cognitive assessments designed for and suited to general practice within the UK and beyond must be a priority for research, taking into account not only administration time but also adaptability for administration by various practitioners, both medical and nursing.

A major research priority should focus upon GP attitude and management of the patient and their family within the context of brief cognitive assessments. It was notable that few GPs mentioned patients at all beyond the targeted questions, and when they did it was in terms of using the tests to reassure patients. The discovery that some GPs would consider knowledge of the patient and patient history to be as important or more important than test performance needs further exploration; is this due to uncertainty with the tests, or prioritising patient management beyond assessment? The common difficulty voiced by GPs when managing patients who present with a number of conditions or issues is also worthy of further research; how does this impact patient testing, and how can the patient journey be improved to aid the testing process within the practical time constraints of the GP consultation? Development of new tests, new testing methods and new technologies such as decision-making algorithms also need assessment and full evaluation using appropriate methodological approaches.

One final topic worth further exploration is the drivers for referral from general practice to specialist services such as psychiatry or gerontology. The results of this survey suggest that the results of brief cognitive assessments have little bearing for many general practitioners when considering further investigation and management decisions. A recent randomised controlled trial by Gill Livingston and colleagues[92] investigating the effectiveness of an intervention to aid prompt referral to memory clinics in the United Kingdom found little effect of the intervention. The authors propose that persistently-low referral rates may have been due to GP concerns around service availability, length of waiting lists for diagnostic services and limited resources for post-diagnostic support[92]. These issues may well impact on GP referral decisions and our survey results suggest this would be a worthwhile area for further investigation.

# Appendix 13 Copy of the survey

## How do GPs choose and use brief cognitive assessments as part of the process for identifying dementia in general practice?

### Page 1

Thank you for following the link to this survey of UK general practitioners (GPs).

I am investigating GP clinical practice and attitudes towards the use of brief cognitive assessment tools (such as the MMSE) as part of the process for identifying dementia in general practice.

This research is part of a wider PhD project looking at how the use of brief cognitive assessments in general practice can contribute to improved dementia diagnosis. If you would like to find out more about this and other projects, please follow the link to our University of Exeter Medical School Test Group webpage: http://medicine.exeter.ac.uk/testgroup/research/dementia/.

**This survey should take around 7 minutes to complete.**

By continuing you consent to taking part in the survey. Completion of the survey is taken as consent for data to be used in the ways described.

All data will be anonymised. Your responses are not traceable to you (see encryption detail below). I will use survey responses to create summary and descriptive statistics about current clinical practice and attitudes towards the use of brief cognitive assessments, and this will contribute to my PhD research. *Please answer all questions as partially completed surveys will not be included in analysis.*

You can withdraw from the survey at any point without giving a reason. I don't believe the survey is likely to cause you discomfort or distress, but if there is any part of it that you'd like to comment on, please contact me at h.a.hunt@exeter.ac.uk.

Thank you for your time -

**Harriet Hunt** | *PhD candidate*

Exeter Test Group | PenCLAHRC | University of Exeter Medical School

**Data protection** Survey data are encrypted and survey responses are sent over an encrypted SSL connection; further information available on the BOS website here:https://www.onlinesurveys.ac.uk/help-support/bos-security/. Data will be stored electronically on University of Exeter Medical School secure servers, with data backup on the University secure cloud-based system.

1 / 5

Page 2

**About you**

How old are you?

- ○ <35
- ○ 35-45
- ○ 46-55
- ○ >55
- ○ Prefer not to say

How many years is it since your first GP appointment?

More info

Please enter the first section of your current practice's postcode (e.g. TQ9, BS16):

More info

Do you have dementia specialists in your practice?

More info

- ○ Yes
- ○ No

If Yes, briefly describe their role below:

*Approximately* how many patients do you assess for possible cognitive impairment or dementia **per month**?

More info

**Brief cognitive assessment tools: familiarity and use**

Which brief cognitive assessment tool(s) do you use **most often**? [you can select more than one]

More info

- ☐ Mini Mental State Examination (MMSE)

2 / 5

- ☐ General Practitioner assessment of Cognition (GPCOG)
- ☐ Mini-Cog
- ☐ Clock Drawing Test (CDT) [as a discrete tool rather than as part of another tool]
- ☐ Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE)
- ☐ None
- ☐ Other

If you selected Other, please specify:

What guides your choice of brief cognitive assessment?

Please select between 1 and 7 answers.

- ☐ Patient factors (e.g. age, level of education, impairments)
- ☐ It's what I've always used
- ☐ Local factors (e.g. availability, surgery preferences)
- ☐ Guidelines (e.g. NICE, Alzheimer's Society)
- ☐ Time taken to administer
- ☐ Cost
- ☐ Other

If you selected Other, please specify:

Within the past 2 years...

| | Have you used this brief cognitive assessment tool **for initial assessment**? | | Have you used this brief cognitive assessment tool **for monitoring decline over time**? | |
|---|---|---|---|---|
| | Yes | No | Yes | No |
| Mini Mental State Examination (MMSE) | ○ | ○ | ○ | ○ |
| General Practitioner assessment of Cognition (GPCOG) | ○ | ○ | ○ | ○ |
| Clock Drawing Test (CDT) | ○ | ○ | ○ | ○ |
| Mini-Cog | ○ | ○ | ○ | ○ |
| Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) | ○ | ○ | ○ | ○ |
| Other [please detail below] | ○ | ○ | ○ | ○ |

If you selected 'Other' please name the brief cognitive assessment below: *Optional*

More info

3 / 5

In your view, what would be an acceptable administration time [in minutes] of a brief cognitive assessment if:

More info

| | | Please comment on your answer *Optional* |
|---|---|---|
| A GP carried out the assessment | | |
| Another member of staff (e.g. a practice nurse) carried out the assessment | | |

**Confidence in practice**

Using the drop-down menu, please select the brief cognitive assessment **you use the most**, and indicate how confident you feel using it on a scale of 1 (*not at all confident*) to 5 (*very confident*).

| | Brief cognitive test | If you selected Other, please specify: | 1 (not at all confident) to 5 (very confident) 1 | 2 | 3 | 4 | 5 | Please would you explain you (e.g. if you feel the need to b confident, what would help w *Optional* |
|---|---|---|---|---|---|---|---|---|
| Interpreting the results of (a) brief cognitive assessment(s)? | Please select | | ○ | ○ | ○ | ○ | ○ | |
| Explaining the results of (a) brief cognitive assessment(s) to patients? | Please select | | ○ | ○ | ○ | ○ | ○ | |

Please add any further comments you have about brief cognitive assessments, particularly in relation to identifying possible dementia *Optional*

4 / 5

## Page 3: Thank you

Thank you for your time, I greatly appreciate your contribution to this research. If you found this survey interesting, please do pass it on to any GP colleagues - the survey is open until midnight on Friday 22nd September 2017.

If you have any comments on this survey or issues raised within it, please do get in touch at h.a.hunt@exeter.ac.uk. Equally if you are interested in the results of the survey or my work more broadly, please contact me at h.a.hunt@exeter.ac.uk

Best wishes,

Harriet Hunt

## Key for selection options

**9.1.a - Brief cognitive test**

MMSE
GPCOG
CDT
Mini-Cog
IQCODE
Other test [please detail]

**9.2.a - Brief cognitive test**

MMSE
GPCOG
CDT
Mini-Cog
IQCODE
Other

5 / 5

## Appendix 14 Ethics application form

**UNIVERSITY OF EXETER MEDICAL SCHOOL**

**RESEARCH ETHICS COMMITTEE**

**APPLICATION FORM**
**FOR**
**RESEARCH ETHICS APPROVAL**

| | |
|---|---|
| **Name of Applicant:** | Harriet Hunt |
| **Project Title:** | General practitioners' clinical practice and attitudes to using brief cognitive assessments to identify dementia |
| **Date:** | 28/06/17 |
| **Version Number:** *(1 for first time applications)* | 1 |
| **Application Number:** (For Ethics Committee use only) | |

# SECTION A: GENERAL

| | | | |
|---|---|---|---|
| 1 Title of the Study: | General practitioners' clinical practice and attitudes to using brief cognitive assessments to identify dementia | | |
| Project Start Date: | 28/06/2017 | Project End Date: | 14/07/2017 |

| | | | | | |
|---|---|---|---|---|---|
| 2 Full name of applicant: Harriet Hunt | | | | | |
| Position Held: | PhD Research at the University of Exeter Medical School | | | | |
| Institution: | University of Exeter | Course Title (if student): | PhD Medical Studies | | |
| Location: | | Exeter | | | |
| Email: | Hh366@exeter.ac.uk | Telephone: | 01392 726074 | Fax: | - |
| Please provide details of any and all other researcher(s) who will work on the research project: (if more than three researchers please extend table as appropriate) | | | | | |
| Name(s): | Professor Chris Hyde | | | | |
| Position Held: | Professor of Public Health & Clinical Epidemiology<br>Primary PhD Supervisor | | | | |
| Location: | Exeter | | | | |
| Contact details (e-mail/ telephone/fax): | C.J.Hyde@exeter.ac.uk | | | | |
| Name(s): | Dr Anne-Marie Boylan | | | | |
| Position Held: | CLAHRC Research Fellow, Nuffield Dept of Primary Care Health Sciences, University of Oxford | | | | |
| Location: | Oxford | | | | |
| Contact details (e-mail/ telephone/fax): | anne-marie.boylan@phc.ox.ac.uk | | | | |
| Name(s): | Dr Nick Cartmell | | | | |
| Position Held: | GP & Primary Care Dementia Adviser to PenCLAHRC | | | | |
| Location: | Ashburton | | | | |
| Contact details (e-mail/ telephone/fax): | nickcartmell@nhs.net | | | | |
| Name(s): | Dr Mark Pearson | | | | |
| Position Held: | Senior Research Fellow in Implementation Science, PenCLAHRC and secondary PhD Supervisor | | | | |
| Location: | Exeter/Plymouth | | | | |
| Contact details (e-mail/ telephone/fax): | Mark.Pearson@exeter.ac.uk | | | | |
| Name(s): | Dr David Llewellyn | | | | |
| Position Held: | Senior Research Fellow in Clinical Epidemiology and secondary PhD Supervisor | | | | |
| Location: | Exeter | | | | |
| Contact details (e-mail/ telephone/fax): | David.Llewellyn@exeter.ac.uk | | | | |
| Name(s): | Professor Willie Hamilton | | | | |
| Position Held: | GP & Professor of Primary Care Diagnostics | | | | |
| Location: | Exeter | | | | |
| Contact details (e-mail/ telephone/fax): | W.Hamilton@exeter.ac.uk | | | | |

| | | | | |
|---|---|---|---|---|
| **3 Is this proposal part of a PhD?** | Yes | ✓ | No | |
| ***If yes, please complete the remainder of this section.*** | | | | |
| Year of Study: | 3 | | | |

| | | | |
|---|---|---|---|
| Name of Primary Supervisor/Director of Studies: | Professor Chris Hyde | Position held: | Professor of Public Health & Clinical Epidemiology |
| Location: | Exeter | | |
| Contact details (email/telephone/fax): | C.J.Hyde@exeter.ac.uk | | |
| Name of Second Supervisor: | Dr Mark Pearson | Position held: | Senior Research Fellow in Implementation Science, PenCLAHRC |
| Location: | Exeter/Plymouth | | |
| Contact details (email/telephone/fax): | Mark.Pearson@exeter.ac.uk | | |
| Name of Third Supervisor: | Dr David Llewellyn | Position held: | Senior Research Fellow in Clinical Epidemiology |
| Location: | Exeter | | |
| Contact details (email/telephone/fax): | David.Llewellyn@exeter.ac.uk | | |

**4 Declaration to be signed by the Applicant or the supervisor in the case of a student:**

I confirm that the research will be undertaken in accordance with the University Ethical Framework, Good Research Practice Policy, and Code of Research Ethics.

I will undertake to report formally to the relevant University Research Ethics Committee for continuing review approval.

I shall ensure that any changes in approved research protocols are reported promptly for approval by the relevant University Ethics committee.

I shall ensure that the research study complies with the appropriate regulations and relevant University of Exeter policies on the use of human material (if applicable) and health and safety.

I shall ensure that any external permissions necessary for the research to be undertaken are obtained prior to the research taking place.

I am satisfied that the research study is compliant with the Data Protection Act 1998, and that necessary arrangements have been, or will be, made with regard to the storage and processing of participants' personal information and generally, to ensure confidentiality of such data supplied and generated in the course of the research.
*(Note: Where relevant, further advice is available from the University of Exeter Medical School [359] Data Protection Officer).*

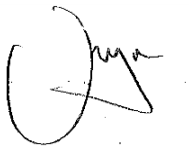
I will ensure that all adverse or unforeseen problems arising from the research project are reported in a timely fashion to the Chair of the relevant University Research Ethics Committee.

I will undertake to provide notification when the study is complete and if it fails to start or is abandoned.

I have met and advised the student on the ethical aspects of the study design and am satisfied that it complies with the current professional (*where relevant),* School and University guidelines.

I have read this application and believe it to be scientifically and ethically sound

**Signature of Applicant:** **Date: 22/06/17**



**Signature of Supervisor:** **Date: 25/06/17**

**Departmental Approval**

- I give my consent for the application to be forwarded to the University of Exeter Medical School Research Ethics Committee with my recommendation that it be approved.
- I confirm that this submission has been appropriately peer reviewed.

**Signature of Head of Research Institute/Centre or Vice Dean (Education) (or approved nominee)**

**Signature:** **Date: 27th June 2017**
**Printed Name: Professor Stuart Logan**

| **5 Name and affiliation of Peer Reviewer(s)** | | | |
|---|---|---|---|
| Name: | Rebecca Whear | Position held: | Research Fellow |
| Institution: | PenCLAHRC/UEMS | | |
| Contact details (email/telephone/fax): | R.S.Whear@exeter.ac.uk<br>01392 726064 | | |

## SECTION B: FUNDING

| **6 If the research is externally funded, what is the source of the funding?** | | | | |
|---|---|---|---|---|
| This research is part of a PhD project funded by PenCLAHRC under the Diagnostics theme | | | | |
| **6.1** What is the value of the grant? n/a | | | | |
| **6.2** Are there any conditions attached to the funding which could have an impact on this application? | | | | |
| **YES** | | **NO** | ✓ | |
| If yes, please specify. | | | | |

## SECTION C: THE RESEARCH

**Background**
In the UK, general practice is often the first place that people will present with concerns about their or a relative's cognition, yet efforts to identify dementia in this setting can often be hampered by non-specific symptoms, multi-morbidities and lack of resources such as consultation time and ability to use complex cognitive assessment tools. Surveys have explored general practitioners' attitudes to dementia identification and diagnosis within primary care, yet the selection and extent to which GPs actually use brief cognitive assessments to identify dementia and refer to memory clinics is unclear.

Factors which affect a tool's utility beyond diagnostic accuracy include: time to administer: suitability for the target population; adaptability for different scenarios; clarity for clinical and patient users; cost; and possible increased value in combination with other tests.

**Aim**

To investigate current clinical practice in the use of brief cognitive assessments to evaluate patients presenting with cognitive concerns, and how that correlates with current research and the established evidence base for the use of such assessments.

**Research methodology**

Through a survey with general practitioners, we will explore:

a) GPs use of brief cognitive assessments as part of their decision making to identify patients likely to have as-yet undiagnosed dementia;
b) The most commonly-used assessment tools; and
c) What, how, and to what extent, does information from the assessment affects GPs' management decisions?

These insights will help build a better picture of how GPs choose and use brief cognitive assessments as part of their process for identifying patients with probable dementia in general practice, to better inform policy makers, test developers, researchers and healthcare professionals and aid primary care in supporting a more effective route to dementia diagnosis.

Before launching, the survey has been tested for flow and technical fluency, and piloted with general practitioners to check sense and facility. This has led to several refinements in the final survey (see attachment).

Quantitative data will be reported via ranked preferences of test and percentage groupings with confidence intervals reported. Univariate descriptive analyses will be used to explore demographic variables and GP attitudes to the use of cognitive assessments in general practice. Qualitative data responses will be analysed thematically to code and develop emerging commonalities, using expertise within our research team to test and refine the analytical framework within which themes will be developed.

**Contribution of research**

This survey is the final component in a programme of doctoral research aiming to address the second part of question 3 from the 2013 James Lind Priority Setting Partnership on Dementia: how can primary care support a more effective route to dementia diagnosis? The James Lind Alliance is an organisation that gathers the views of patients, clinicians, researchers and experts on a specific topic and through a series of workshops identifies a core set of research priorities for that topic. Our intention is that this survey will provide initial data to contribute to a larger research project exploring further how practitioners and patients communicate and understand diagnostic information within the general practice setting.

**Justification of benefit**

Whilst the timely identification of dementia and referral to specialist services is important to improve the chances of living well with dementia for as long as possible, general practice assessment needs to be balanced with implications of the increased likelihood of false positives and moderate test performance of brief cognitive assessments within this setting (Pond et al., 2013; Jacova et al., 2007; Yokomizo, Simon & de Campos Bottino, 2014). Recommendations for operationalizing the assessment of cognitive performance and detection of dementia have been created as part of national dementia strategies by several countries including the United States (GSA, 2015), the UK (DH, 2016), and Australia (GAC, 2016), and they all discuss brief cognitive assessment tools in detail, as well as highlighting the value of employing appropriate tools as part of an iterative diagnostic process. These strategies commonly acknowledge many of the challenges in finding the most suitable tool for the primary care setting. What they lack are clear recommendations and guidance on how the best tool should be chosen by the primary care practitioner, and when tools should – and should not – be used as part of the diagnostic routine.

This survey will contribute towards a PhD.

**References**

Pond, C.D., et al., Predictors of agreement between general practitioner detection of dementia and the revised Cambridge Cognitive Assessment (CAMCOG-R). International psychogeriatrics, 2013. **25**(10): p. 1639-1647.
Jacova, C., et al., Neuropsychological testing and assessment for dementia. Alzheimer's & Dementia, 2007. **3**(4): p. 299-317.
Yokomizo, J.E., S.S. Simon, and C.M. de Campos Bottino, Cognitive screening for dementia in primary care: a systematic review. . International Psychogeriatrics, 2014. **26**((11)): p. 1783-1804.
GSA, The Gerontological Society of America Workgroup on Cognitive Impairment Detection and Earlier Diagnosis: Report and recommendations, GSA, Editor. 2015.
DoH (Department of Health) Prime Minister's Challenge on Dementia 2020. 2016.
GAC (Guideline Adaptation Committee), Clinical Practice Guidelines and Principles of Care for People with Dementia 2016, Guideline Adaptation Committee: Sydney.

Include any questionnaires, psychological tests, etc. at the end of your application.

**8 Location of study**

**8.1** Where will the study take place?

In the UK

**8.2** If the study is to be carried out overseas, what steps have been taken to secure research and ethical permission in the country of study? (Please attach evidence of approval if available.)

n/a

**9 Multi-centre and off-campus studies**

**If this is a multi-centre or off-campus study, please answer the appropriate questions below; otherwise, go to Question 11.**

**9.1** Does this project involve a consortium (other research partner organisations)?

| **YES** | | **NO** | | |
|---|---|---|---|---|

If yes, please complete the details below in Question 9.2.

**9.2** Who has overall responsibility for the study?

Please provide details of the contractual agreement between UEMS and the other organisation(s).

**9.3** Is this an off-campus study?

| **YES** | | **NO** | | |
|---|---|---|---|---|

If yes, please provide signed, written permission from an appropriate level of management within the relevant organisation(s).

**10 Has approval been sought from other Ethics Committees and LRECs?**

| **YES** | | **NO** | | |
|---|---|---|---|---|

**11 Who will have overall control of the data generated?**

Harriet Hunt at the University of Exeter

**12 How do you propose to disseminate the results of your research?**

Peer reviewed journal publications, PhD thesis, academic conferences and via workshops and focus groups for follow-up work (subject to post-doctoral funding).

**13 METHODS AND PROCEDURES**

| Describe the nature of the task required of participants and the various precautionary measures to be taken to avoid harm or discomfort if appropriate. If the study is likely to cause discomfort or distress to **subjects** [?], estimate the degree and likelihood of discomfort or distress. |
|---|
| Participants are to complete a survey exploring clinical practice and attitudes towards using brief cognitive assessment to identify dementia in general practice. This will be via an electronic survey taking around 7 minutes to complete. Participants will be asked if they consent to taking part in the survey, and told they can withdraw at any point without giving a reason. It is not anticipated that the survey is likely to cause discomfort or distress to participants.<br>I have included a copy of the survey form at the end of this application.<br>*N.B. I note the use of 'subjects' in the question above (my **bold** and boxed question mark) and query if it is appropriate, or would be better replaced with 'participants'.* |

**13.1** Does the study include any of the following interventions / invasive procedures?

| | YES | NO | | YES | NO |
|---|---|---|---|---|---|
| Participant-observation / non participant-observation | ☐ | ☒ | Self-completion questionnaires | ☒ | ☐ |
| Interviews | ☐ | ☒ | Video / audio recording | ☐ | ☒ |
| Focus Groups | ☐ | ☒ | Administration of substance / drug (e.g. caffeine / doubly labeled water etc) | ☐ | ☒ |
| Physical examination | ☐ | ☒ | Manipulation of diet | ☐ | ☒ |
| Arterial puncture* | ☐ | ☒ | Venepuncture* | ☐ | ☒ |
| Urine sample* | ☐ | ☒ | Fingertip blood sample* | ☐ | ☒ |
| | | | Saliva sample* | ☐ | ☒ |
| Body Imaging (e.g. MRI, DEXA, X-rays) | ☐ | ☒ | | | |
| * if yes, will samples be retained for subsequent testing for factors other than described in this proposal? | | | | ☐ | ☐ |
| If yes, will samples be anonymised? | | | | ☐ | ☐ |

**If you are using human tissue in your project, you must complete section E.**

| **14 Products and devices** | | | | |
|---|---|---|---|---|
| **14.1** Does the research involve the testing of a product or device? | | | | |
| **YES** | | **NO** | ✓ | |
| **14.2** If this research involves a drug, is it being used in accordance with its licensed uses? | | | | |
| **YES** | | **NO** | ✓ | |

## d) SECTION D: THE PARTICIPANTS

| | |
|---|---|
| *For the purposes of this section, "participants" include human subjects, their data, their organs and/or tissues.* For participants to be recruited to the research, please state: | |
| **15 Number of participants:** | Approximately 200 |
| **16 If data are to be collected on different sites, please state the number of participants at each site: n/a** | |
| **17 How have you arrived at this number? Please state proposed inclusion/exclusion criteria. If appropriate has the protocol been reviewed by a Statistician?** | |
| We are employing an opportunistic sampling strategy to conduct an exploratory survey, and are therefore unable to give precise numbers. All UK general practitioners (including those employed part time, full time and locums) are eligible to complete the survey. We will distribute the survey using the Contact, Help, Advice and Information Network (CHAIN - http://www.chain-network.org.uk) which has approximately 170 general practitioners on the mailing list. We will also use our professional networks such as the South West Clinical Research Network (SWCRN) and Exeter Collaboration for Academic Primary Care (APEx) research group to publicise the survey and to post on a GP-specific Facebook group. In addition, we will use Twitter to promote the survey, using the hashtag #gpnews and requesting retweets from the Royal College of General Practitioners (@RCGP), Devon Local Medical Committee (@Devon_LMC) and practitioner magazines such as GPOnline.com (@GPonlinenews) and PulseToday (@pulsetoday). I also plan to place a note in the Devon Local Medical Committee newsletter to promote the survey.<br>Using CHAIN and our professional networks as outlined, we anticipate reaching ~200 people and anticipate a response rate of around 50%, based on previous experiences of colleagues who have undertaken similar survey work with similar target groups. This should allow approximately 100 responses to the survey.<br>I have discussed this protocol with Obi Ukoumunne, Associate Professor in Medical Statistics at the University of Exeter Medical School. | |
| **18 Age group or range *(e.g., under 60s)*:** | 18 – 80 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **18.1 Sex:** | Male | | | Female | | |
| **19. Is this a single sex study?** | | | | | | |
| | **YES** | | **NO** | ✓ | | |
| If yes, please justify the reason(s) for gender selection | | | | | | |
| | | | | | | |

*While some studies explicitly focus on gender specific experiences, care should be taken to ensure that women or men are not unnecessarily excluded from participating in research.*

| | | | | | |
|---|---|---|---|---|---|
| **20 Do participants belong to any of the following vulnerable groups?** | | | | | |
| Children: | **YES** | | **NO** | ✓ | |
| Participants unable to give informed consent in their own right *(e.g., people with learning difficulty)*: | | | | | |
| | **YES** | | **NO** | ✓ | |
| Other vulnerable groups (please specify) | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | **YES** | | **NO** | ✓ | |

| |
|---|
| *Care will need to be taken to formulate inclusion/exclusion criteria that clearly justify why certain individuals are to be excluded, to avoid giving the impression of unnecessary discrimination. On the other hand, the need to conduct research in "special" or "vulnerable" groups should be justified and it needs generally to be shown that the data required could not be obtained from any other class of participant.* |

If the answer to any of the above is yes, please complete Questions 21 to 25; otherwise proceed to Question 26.

| |
|---|
| **21 Please explain why it is necessary to conduct the research in such vulnerable participants and whether required data could be obtained by any other means.** |
| **22 Please state what special or additional arrangements have been made to deal with issues of consent and the procedures to safeguard the interests of such participants.** |
| **23 Please describe the procedures used to ensure children (i.e., persons under 18 years) are able to provide consent/assent to participation.** |
| **24 If appropriate, please state whether and how parental consent, or the consent of the legal guardian and/or order/declaration of the court, will be sought in relation to the participation of children in the research.** |
| **25 If the participant is unable to consent in their own right, will you seek the prior approval of an informed independent adult and any other person or body to the inclusion of the participant in the research?** |

| | | | | | |
|---|---|---|---|---|---|
| | **YES** | | **NO** | | |

| |
|---|
| State precisely what arrangements will be put in place. |
| |

| **Recruitment and Selection** |
|---|
| *The Research Ethics Committee will need to be satisfied with the effectiveness and propriety of recruitment and selection procedures given the participant involved, e.g., that the participant will not feel in any way obliged to take part, that advertisements do not appear to offer inducements. The Committee will be particularly interested in cases where a participant's relationship with the investigator could raise issues about the voluntary status or motive of the participant's involvement in the research (e.g., students).* |
| **26 How will the participants in the study be selected, approached and recruited (please indicate the inclusion and exclusion criteria)?** |
| We are employing an opportunistic sampling strategy to conduct an exploratory survey. All general practitioners (including those employed part time, full time and locums) are eligible to complete the survey. We will distribute the survey using the Contact, Help, Advice and Information Network (CHAIN - http://www.chain-network.org.uk) which has approximately 170 general practitioners on the mailing list. We will also use our professional networks such We will also use our professional networks such as the South West Clinical Research Network (SWCRN) and the Exeter Collaboration for Academic Primary Care (APEx) research group to publicise the survey and to post on a GP-specific Facebook group. We will also use Twitter to promote the survey, using the hashtag #gpnews and requesting retweets from the Royal College of General Practitioners (@RCGP), Devon Local Medical Committee (@Devon_LMC) and practitioner magazines such as GPOnline.com (@GPonlinenews) and PulseToday (@pulsetoday). |

| | | | | | |
|---|---|---|---|---|---|
| Potential participants will follow a weblink to the online survey, and once they have read the information and given informed consent they will be guided through the survey via the BOS portal. | | | | | |
| *If you are proposing to advertise, please include a copy of the advert to be used at the end of your application.* | | | | | |
| **27 Where are you recruiting the participants?** | | | | | |
| We will distribute the survey using the Contact, Help, Advice and Information Network (CHAIN - http://www.chain-network.org.uk) which has approximately 170 general practitioners on the mailing list. We will also use our professional networks such as the South West Clinical Research Network (SWCRN) and Exeter Collaboration for Academic Primary Care (APEx) research group to publicise the survey and to post on a GP-specific Facebook group. We will also use Twitter to promote the survey, using the hashtag #gpnews and requesting retweets from the Royal College of General Practitioners (@RCGP), Devon Local Medical Committee (@Devon_LMC) and practitioner magazines such as GPOnline.com (@GPonlinenews) and PulseToday (@pulsetoday). | | | | | |
| **28 Relationship of participant to investigator:** | | | | | |
| **29 Will the participants take part on a fully voluntary basis?** | | | | | |
| | **YES** | ✓ | **NO** | | |
| **30 Will students (e.g. PCMD, UEMS, other Schools or Colleges) be involved as participants in the research project?** | | | | | |
| | **YES** | | **NO** | ✓ | |
| If yes, please provide full details. | | | | | |
| | | | | | |
| **31 Will payments or other inducements be made to participants?** | | | | | |
| | **YES** | | **NO** | ✓ | |
| If yes, give amounts, type and purpose. | | | | | |
| | | | | | |
| **Information to Participants and Consent**<br>*If your study involves the collection and storage of human samples, please refer to the University Human Tissue Act Management Handbook and follow the guidelines for obtaining informed consent.* | | | | | |
| **32 Will participants be informed of the purpose of the research?** | | | | | |
| | **YES** | ✓ | **NO** | | |
| If no, please explain why. | | | | | |
| | | | | | |
| **33 Will the participants be given a written information sheet?** | | | | | |
| | **YES** | | **NO** | ✓ | |
| If no, please explain why and delete Appendix 1. | | | | | |
| This information is given as the front page of the survey with a forced choice tick box for consent – 'yes/no' at the bottom on the page. If the 'no' tick box is selected, the participant is directed to an information page with a message to clarify they have not consented and thanking them for their time. | | | | | |
| **34 Will written consent be obtained?** | | | | | |
| | **YES** | ✓ | **NO** | | |
| Please see the survey attached to this application. | | | | | |
| **35 Where potential participants will/may suffer from any difficulties of communication, state the methods to be employed both to present information to** | | | | | |

| |
|---|
| **the participants and achieve consent.** *If written, please include a copy at the end of your application.* |
| As distribution will be via email and Twitter, and response will be via an online survey, only respondents with the means to access these modes will be eligible and this will be acknowledged within the survey. |
| **36 Ensure that the Information Sheet includes details of the participants' right to withdraw from the study at any time without penalty.** |
| Where relevant (should incidental significant findings emerge during the course of a study) |
| **36.1** Will any information be given to the participants' GP (if deemed necessary)? |

| | | | | | |
|---|---|---|---|---|---|
| | **YES** | | **NO** | ✓ | |

| |
|---|
| **36.2** Have the participants consented to having their GP informed? |

| | | | | | |
|---|---|---|---|---|---|
| | **YES** | | **NO** | ✓ | |

| |
|---|
| **37 Please state what measures will be taken to protect the confidentiality of the participant's data (i.e., arising out of the research and contained in personal data).** |
| All data will be anonymised and potential identifiers (such as age, years in practice) will be banded into ranges. No names will be recorded unless personal feedback is asked for by respondents via an email link. |
| **38 How will the data be stored during the life of the project?** |
| The survey will be designed and distributed using Bristol Online Surveys (BOS - www.onlinesurveys.ac.uk). This is a stable and robust academic and research survey tool developed by the University of Bristol. All data is currently stored securely in the UK via the University of Bristol servers, although we have recently been notified that this service is transferring from the University of Bristol to Jisc (see letter in Appendix). This will not change storage details except that after 30th June 2017 the raw survey data will be hosted by Jisc in a secure cloud based infrastructure with servers located in the Republic of Ireland, ensuring this data is held and processed in compliance with Data Protection legislation.<br>Survey data is encrypted and survey responses are sent over an encrypted SSL connection. BOS user passwords are encrypted using PBKDF2 with a SHA256 hash and a random salt.<br>Analytical data will be stored securely on the University of Exeter secure servers in line with University Guidelines (see below). |
| **39 University of Exeter Guidelines state that primary data generated in the course of research must be kept securely in paper or electronic format, as appropriate and held normally for a period of five years (or as required by the funding body) after the completion of a research project.**<br>http://www.exeter.ac.uk/research/toolkit/throughout/ethics/goodpractice/<br>**Please provide details of how data will be stored, how long the data will be retained following completion of the study and how the data will be disposed of once this period has ended** |
| Data will be stored electronically on University of Exeter Medical School secure servers, with data backup on the University secure cloud-based system. Data relating to the survey (e.g. downloaded .csv files, all analytical programmes used and reported produced from these data) will be retained for 5 years until 31/08/2022 with Professor Chris Hyde responsible for the data until that point and responsible for disposing of the data after that date. Disposal will be in line with current technology through electronic deletion of files and copies associated with them. |
| **40 Who will be ultimately responsible for data storage and disposal for this project?** |
| Harriet Hunt will be ultimately responsible for the data storage and disposal after 31/08/2022. |

| **41 How will participants be informed of the results of the study if they so wish?** |
|---|
| Participants will be offered the opportunity to feedback on the survey and be kept up to date with the survey results by email. They will also have the opportunity to see these results in the wider context of my PhD thesis if that is of any interest to them! |

| **42 Risk to research participants** | | | | | |
|---|---|---|---|---|---|
| **42.1** do you think there are any ethical problems or special considerations/hazards with the proposed Study? If so, please describe | | | | | |
| **No** | | | | | |
| **43 Does your proposed study require a Health and Safety risk assessment and if so, has this been carried out?** | | | | | |
| **YES** | | **NO** | ✓ | | |
| **44 Are there any potential conflicts of interest arising from the project, deriving from relationships with collaborators/sponsors/participants/interest groups?** | | | | | |
| | **YES** | | **NO** | ✓ | |
| Please disclose all relevant personal and commercial interests. | | | | | |
| NC and WH are both practising GPs so have personal and professional interests from being employed within the profession.<br>NC also provides consultancy services to a private company in Plymouth, Re:Cognition Health, which involves recruiting people to new drug Alzheimer's Disease clinical trials and monitoring them during such trials. He is paid for this work on a sessional rather than per-patient basis. | | | | | |

UNIVERSITY OF EXETER | MEDICAL SCHOOL

University of Exeter Medical School Research Ethics Committee

**Reviewer Form**

| | |
|---|---|
| **Name of Reviewer:** | Rebecca Whear |
| **Employing Organisation:** | PenCLAHRC/UEMS |
| **Qualifications and area of expertise:** | Health services researcher |
| **Details of any potential conflict of interest:** | N/A |
| **Name of Researcher:** | Harriet Hunt |
| **Project Title:** | General practitioners' clinical practice and attitudes to using brief cognitive assessments to identify dementia |

| | **Yes** | **No** | **N/A** |
|---|---|---|---|
| Is there a clear research question/aim? | ☒ | ☐ | ☐ |
| Are the methods of data collection adequately outlined? | ☒ | ☐ | ☐ |

| | | | |
|---|---|---|---|
| Are the methods of data collection appropriate? | ☒ | ☐ | ☐ |
| Are the methods of data collection adequately described? | ☒ | ☐ | ☐ |
| Are the methods of data analysis appropriate? | ☒ | ☐ | ☐ |
| Have ethical issues been addressed appropriately? | ☒ | ☐ | ☐ |

Please grade each feature (where appropriate) from excellent to very poor:
Evaluation Scale: (5) Excellent (4) Very Good (3) Good (2) Fair (1) Poor

| | |
|---|---|
| Originality | Excellent |
| Reliability | Very Good |
| Importance | Excellent |

What is your overall assessment of the **quality of the study**?. (please continue overleaf)

The overall quality of this study is excellent with clear thought as to how the questionnaire might be received and the time available to practitioners to complete it. This is good evidence of secure data collection and storage although how the two marry together could be better described (see below). I could not identify any ethical issues that would be of concern to the participants outside of what has already been covered by the researcher. However, there are a few points I would like to see clarified as below.

What specific improvements would you like to see the applicant make in relation to the **quality of the study**?.

I am not sure that the current survey captures the third research objective 'to what extent, does information from the assessment affects GPs' management decisions?' as there are no questions related to what GP's do with the information they gain or how it impacts further decisions. Could this be clarified or improved.

*HH response: my intention is that this objective is addressed more obliquely through 'other' options and comments sections.*
*Q.7 "Have you ever used this brief cognitive test for initial assessment?" and "Have you ever used this brief cognitive test for monitoring decline over time?" are designed to surface the different management decisions GPs commonly make with these tests, and Q.9 "…please select the brief cognitive assessment you use the most, and indicate how confident you feel on a scale of 1 (not at all confident) to 5 (very confident)" should allow me to link previous responses to show confidence in use of specific tests used for purposes selected in Q.7.*

*This should indicate – where the information is provided – level of confidence in the tests used for the purposes indicated.*

*You're quite right that the impact on further decision making is not explicitly addressed, although I anticipate there will be some open question responses which touch on this in Q.10 "Please add any further comments you have about brief cognitive assessments, particularly in relation to identifying possible dementia".*

In the dissemination would a seminar for local GPs/clinics that got involved in the survey be beneficial?

*HH - I'd love to do this sort of activity but don't have time or funds within the bounds of the research project. I have thought long and hard about the balance between this being an entirely anonymous survey and the consequent benefits of attracting comments 'with no strings attached' versus asking people for follow up activities. And as my supervisors asked*

*"when will you have time for that?" I've taken a pragmatic course whilst also offering to share results and discuss findings at the end of the survey.*

Clarification as to how the data will be collected by JISC and transferred and stored by Exeter would be beneficial.

*HH – thank you – I've further clarified sections 38 and 39 to address this point.*

Would the researcher consider it beneficial to ask a question about how long the GP thinks it takes them to use the tool they are using in practice in order to further inform their use/barriers to use in practice.

*HH – this is an insightful suggestion as administration time is highly contentious with large ranges reported amongst the evidence and very few empirical measurements of time reported. However this is also highly subjective and varies depending on disease severity. I'm not convinced this data from the survey would add anything further to current evidence, which is similarly based on anecdote – so to keep it short I'll leave this out.*

Lastly, are there any potential **ethical issues/risks** you would like to bring to the attention of the Committee?

None

**Signed:**
*(Electronic signature required)*
**Date: 27/06/17**

# e) Comments from the Chair 28/06/17 and actions taken

| Chair's comments | HH responses |
|---|---|
| I can't see the "advert" or email text that you will be sending through the various networks to recruit people. Can you send a copy of this as an appendix to the application please? | I've added a copy of this to Appendix 2 of the application. |
| On this and the front of the survey, you may want to include - information that tells them no personal data will be collected, what you are interested in collecting, and what will happen to the data they submit (use to calculate descriptive statistics about current practice for example). | I've added the suggested information highlighted in the copy of the updated survey in Appendix 3. |
| Will you use any of the free text they put in the other boxes? How? Also this may need to be anonymised before use if they give, for example, information like the name of practice, in the answer. | I will use the free text data provided for narrative description within the final research paper and thesis chapter; for example:<br>Q.4 asks about dementia specialists in the GP's practice. Where people provide information on dementia specialists, I hope to comment upon the number of respondents with dementia specialists and different characteristics (e.g. 90% of respondents mention dementia nurses, lack of resources noted by 20/113 respondents);<br>Q.6 asks for 'other' brief cognitive assessment tools – there are around 15 other *relatively* uncommon tests I haven't listed, but these vary in popularity by geographical cluster (e.g. dementia champions in an area promote successfully) or time (e.g. Alzheimer's Society released guidelines ~6 months ago which recommended specific tools). I'd simply like to capture as many tools that are actually used without listing all of them which risks putting off the majority.<br>Q.7 asks for factors which guide the choice of test. Here I have listed factors which commonly appear in the research literature, but I'm trying to surface factors currently overlooked or under recognised and hope this may be a way of doing it. Again, this data will be reported narratively with comment made on common/contrasting issues, themes and language.<br>Q.8 'other' option covers similar issues as Q.6 in terms of allowing inclusion of less common tools.<br>Q.11 asks for 'any further comments' and whilst I can see this may be a hostage to fortune in its breadth, much of the research literature is deductive in terms of the questions asked about clinical use of tools. I wanted to |

| | |
|---|---|
| | allow space for issues not yet surfaced to feature, plus any partisan or sceptical views to be voiced (anecdotally it is likely that some responders will express views on clinical futility or cynicism around the benefit of tool use, but I didn't want to direct this too firmly to the exclusion of other views).<br><br>Data will be 'cleaned' before analysis so that any identifiable information (e.g. Practice name) will be removed by the lead reviewer before the data is shared more widely within the research team. |
| You might think about the consent process. For an online survey, we often consider that consent is given by the survey being undertaken, rather than having an explicit consent box. You can make a statement saying that completion of the survey is taken as consent for data to be used in the ways you describe, on the front page of the online survey. In addition, you need to think about what you will do with partially completed surveys – both from an ethics point of view – how many questions count as consent to do the survey? (this may happen even if they tick a box that says they consent to take part!) And a data perspective – what will you do with part completed surveys – will you include in the analysis? Again, whatever you decide can go on the front page of the survey. | Thank you, this is helpful.<br><br>I have removed the need to explicitly consent by ticking a box, and replaced this with a statement as suggested. This should improve flow and make the survey easier/quicker to complete.<br><br>Partially completed surveys will be included if a threshold of 10 out of 11 questions are completed.<br>In my view we may reasonably suppose that one response may be missed accidentally or deliberately, but the positive intent is there to complete the survey. Fewer responses will not be included in analysis, and I have added a line to this effect ("Please answer all questions as partially completed surveys will not be included in analysis") to the opening instructions. Non-eligible survey responses will be kept separately from those to be analysed, but handled the same in terms of anonymity and destruction after 5 years.<br><br>Hopefully partial completion will be minimal for several reasons:<br>Most questions are forced choice and alert the responder if they have left an answer blank;<br>The survey is designed to be short (5-7 minutes, N=6) to reduce fatigue and drop out;<br>There is no facility to complete the survey later. |
| In terms of data storage, we usually suggest that the supervisor is responsible for destruction as 5 years is beyond the end of your PhD. | I have changed this to reflect your advice, i.e. that Professor Chris Hyde will be responsible for destruction (and have notified him of this change). |

f)

# g) Email/contact text for survey launch

## For individual GPs who have agreed to forward this on to their networks:

Dear Dr X,

**Short survey: how do GPs choose and use brief cognitive assessments as part of the process for identifying dementia in general practice?**

Thank you for agreeing/offering [*delete as appropriate*] to share this survey with GP colleagues.

The survey takes around 7 minutes to complete, and no personal data will be collected. I am interested in GPs' views on the use of brief cognitive assessment tools (such as the MMSE) as part of the process for identifying dementia in general practice. The survey can be found here: [*link to the survey*].

I will use survey responses to create summary and descriptive statistics about current clinical practice and attitudes towards the use brief cognitive assessments, and this will contribute to my PhD research at the University of Exeter Medical School. More information on the broader research can be found here:
http://medicine.exeter.ac.uk/testgroup/research/dementia/

If you have any questions or are interested in this work please feel free to contact me at the address below, or my supervisor Professor Chris Hyde at C.J.Hyde@exeter.ac.uk.

The survey closes at midnight on Friday 14th July, so please do encourage others to share this with their GP colleagues.

Many thanks for your help,

**Harriet Hunt**
PhD candidate – PenCLAHRC Diagnostics theme
University of Exeter Medical School
South Cloisters | St Luke's Campus | Exeter | EX1 2LU
01392 726074
h.a.hunt@exeter.ac.uk
http://medicine.exeter.ac.uk/testgroup/
Twitter: @HarrietAHunt

## For circulating to CHAIN:

*For the CHAIN team: Please would you share this with your 'dementia' special interest group?*

**How do GPs choose and use brief cognitive assessments as part of the process for identifying dementia in general practice?**

I am conducting a brief (*7 minutes*) survey of general practitioners (GPs) on the topic of brief cognitive assessment use for identifying dementia. I would be grateful if any GPs could take part in the survey here: [*survey link*] or if people could share with GP colleagues who may be interested.

No personal data will be collected. I am interested in GPs' views on the use of brief cognitive assessment tools (such as the MMSE) as part of the process for identifying dementia in general practice.

I will use survey responses to create summary and descriptive statistics about current clinical practice and attitudes towards the use brief cognitive assessments, and this will contribute to my PhD research at the University of Exeter Medical School. More information on the broader research can be found here:
http://medicine.exeter.ac.uk/testgroup/research/dementia/

The survey closes at midnight on Friday 14th July, so please do encourage others to share this with their GP colleagues.

Many thanks for your help,

**Harriet Hunt**
PhD candidate – PenCLAHRC Diagnostics theme
University of Exeter Medical School
South Cloisters | St Luke's Campus | Exeter | EX1 2LU
01392 726074
h.a.hunt@exeter.ac.uk
http://medicine.exeter.ac.uk/testgroup/
Twitter: @HarrietAHunt

## For the SW Clinical Research Network who have agreed to place this in their newsletter:

**How do GPs choose and use brief cognitive assessments (such as the MMSE) as part of the process for identifying dementia in general practice?**

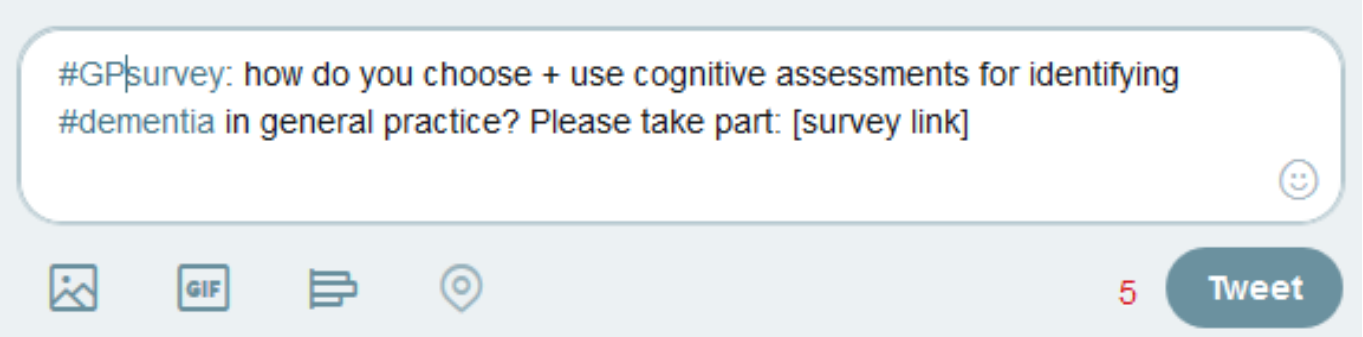
Harriet Hunt, a University of Exeter Medical School PhD candidate, is conducting a brief (*7 minutes*) survey of general practitioners (GPs) on the topic of brief cognitive assessment use for identifying dementia. She would love to hear from any South West GPs who can take part in the survey here: [*survey link*]

No personal data will be collected, and survey responses will be used to create summary and descriptive statistics about current clinical practice and attitudes towards the use brief cognitive assessments. This will contribute to Harriet's PhD research at the University of Exeter Medical School. More information can be found here: http://medicine.exeter.ac.uk/testgroup/research/dementia/ She can be contacted for more information by email: h.a.hunt@exeter.ac.uk

The survey closes at midnight on Friday 14th July, so do encourage others to share this with their GP colleagues.

## For Twitter [140 characters]:

Variants on (i.e. adapted and retweeted every couple of days depending on responses):




Requesting RTs from the Royal College of General Practitioners (@RCGP), Devon Local Medical Committee (@Devon_LMC) and practitioner magazines such as GPOnline.com (@GPonlinenews) and PulseToday (@pulsetoday) plus influencers @Maureenrcgp (former Chair RCGP), @HelenRCGP (current Chair RCGP), @ClareGerada, @muirgray and @trishgreenhalgh.

# Appendix 15 University of Exeter Medical School Ethics Committee Certifical of Approval

UNIVERSITY OF EXETER | MEDICAL SCHOOL

**University of Exeter Medical School**
**Research Ethics Committee**

**Certificate of Ethical Approval**

**Research Institute/Centre:** Institute of Health Services

**Title of Project:** General Practitioners' clinical practice and attitudes to using brief cognitive assessments to identify dementia

**Name(s) of Project Research Team member(s):** Harriet Hunt, Professor Chris Hyde, Dr Anne-Marie Boylan, Dr Nick Cartmell, Dr Mark Pearson, Dr David Llewellyn & Professor Willie Hamilton

**Project Contact Point:** Harriet Hunt

**This project has been approved for the period**

**From:** 30 June 2017

**To:** 31 December 2017

**University of Exeter Medical School**
**Research Ethics Committee approval reference**: Jun17/D/131

**Signature:** RGarside

**Date:** 30 June 2017

**Name of Chair:**
**Ruth Garside, PhD**

Your attention is drawn of the attached paper "Guidance for Researchers when Ethics Committee approval is given", which reminds the researcher of information that needs to be observed when Ethics Committee approval is given.

Application Reference Number 17/06/131

## Appendix 16 Coding and thematic framework summarising key themes

| Codes | Issues discussed | Themes identified |
|---|---|---|
| Usefulness of the test | "Need the best tool available" [Q8]<br>"Seems to be a very good discriminatory test" [Q5]<br>"I don't find this tool particularly discriminating…However, it seems to do the job for reassuring those with no problems and ensuring onward investigation for those in which there are concerns within the history" [Q9]<br>"language can be an issue" [Q9]<br>"it is such a bad scale it can be skewed by anxiety/ bad day etc. and it doesn't pick up subtle issues" [Q9]<br>"fewer false negatives" [Q9]<br>"although I'm also looking at verbal fluency, functional changes" [Q9]<br>"Although useful if a low score to say – needs referral for further testing- looks like a possible memory issue" [Q9]<br>"they're no substitute for a good history" [Q10]<br>"…but if I have enough clinical suspicion, I will refer regardless of high score" [Q10]<br>None of them include questions about functional ability as opposed to cognitive ability" [Q10]<br>"Not always accurate" [Q10]<br>"clock drawing often really helpful" [Q10]<br>"…the quicker ones are quicker but less useful" [Q10]<br>"Difficult to assess how much anxiety and depression affects the results" [Q10]<br>"Would be helpful to have an idea of how to assess premorbid level of functioning for both ends of intellectual spectrum. And how that might affect results of simple tests." [Q10]<br>"It would be useful to have a score system with comments e.g. 0 - 10, 10 - 20, 20 - 30 etc" [Q10]<br>"No perfect one!" [Q10]<br>"I find them a very crude tool and as a GP I don't believe that I am the best placed person to do it." [Q10] | The best test<br>Good/bad test<br>Functional "does the job"<br>Assistive "helpful" "useful" or flawed "crude" "skewed"<br>Not discriminating<br>Not accurate<br>Imperfect |
| Speed of the test/timing | "in the past it was there as a quick tool until MMSE went" [Q5] | Quick tool<br>Brief<br>In context |

| | | |
|---|---|---|
| | "Needs to be 'brief' but for cognitive decline 2 minutes I think should be fine recognising that the brief interventions for smoking for example are different in context." [Q8]<br>"In the context of a total ten minute appointment" [Q8]<br>[Patients] "need time to digest what's going on. " [Q8]<br>"for a GP, you need an extra 10-15 minutes on top of the assessment to deal with this" [Q8]<br>"if you only have 10 mins, the assessment is the thing that gets squeezed" [Q8]<br>"I think a brief assessment tool should be just that, brief" [Q8]<br>"Needs to be quick" [Q8]<br>"They feel like a huge time burden, the quicker ones are quicker but less useful" [Q10]<br>"They need to be short in time to facilitate a realistic assessment in general practice" [Q10] | Competition for time<br>Time burden<br>Trade-off between speed and accuracy |
| Efficiency/ systemic factors | Clinical software - in the past it was there as a quick tool until MMSE went" [Q5]<br>"GPCOG needed for memory clinic referral" [Q5]<br>"Available on EMIS system" [Q5]<br>"I have personalised a version of MMSE and embedded it within EMIS… This makes it easy to administer." [Q5]<br>"1. I remember its name 2. I can find it on the computer" [Q5]<br>"Needs to be 'brief'" [Q8]<br>"I'll sometimes arrange a double appointment" [Q8]<br>"I would make time" [Q8]<br>"less than 10 mins ideally" [Q8]<br>"if only to do the assessment, 15 minutes is ideal" [Q8]<br>"Needs to be consistent within the practice" [Q8]<br>"Needs to be fitted into surgery time" [Q8]<br>"More time to spend in discussion" [Q8]<br>"if funded" [Q8]<br>"If booked as a separate appointment" [Q8]<br>"we have 10 minute appointments" [Q8] | Software as facilitator<br>Software as barrier<br>Tests as access points<br>Personalisation<br>Personal agency<br>Working within system constraints<br>Ideal versus reality<br>Consistency<br>Culture within the general practice<br>GP time is precious<br>Resourcing pressures<br>Technology absolves GP of some responsibility (e.g. remembering scoring)<br>Pressure<br>Process over judgment |

| | | |
|---|---|---|
| | "We are time poor and offering double or review appointments means we are then under more pressure for appts from other patients" [Q8]<br>"gps time prob more productively spent doing something else and anyone trained could do" [Q8]<br>"nurse appts are not resourced" [Q8]<br>"These tend to be long consultatins without the test" [Q8]<br>"lots more to do in the 10 minute consultation" [Q8]<br>"assuming a 10 minute appointment, if a longer appointment the test could take longer" [Q8]<br>"done more slowly and accurately by the nurse" [Q8]<br>"I always have to re-google the normal range scores" [Q9]<br>"its the one I know" [Q9]<br>"if clincial picture is one of worried well but not if more complex picture /sutle problem or high IQ" [Q9]<br>"More difficult if scores between 25 – 29" [Q9]<br>"Familiarity" [Q9]<br>"It's written on the screen... would I remember scoring? Not a chance but I don't need to." [Q9]<br>"They need to be short in time to facilitate a realistic assessment in general practice" [Q10]<br>"Stopped using MMSE when charges introduced for using it" [Q10]<br>"They are not 'brief'. " [Q10]<br>"Can be very time consuming. Also MMSE now copyrighted..." [Q10]<br>"We use 6-CIT as per local guidance and it's part of the referral form. It's complex to score but embedded in SystmOne (computer system used by GP, psych, district nurses etc) so quick to score. More difficult on a home visit" [Q10]<br>"Pressure on time and appointments drives lots of GP decision making" [Q10]<br>"I am rarely surprised by results & do it since I am supposed to" [Q10]<br>"as a GP I don't believe that I am the best placed person to do it" [Q10]<br>"as a single element of testing the serial months backwards is a fantastic discriminator" [Q10] | |

| | | |
|---|---|---|
| Context | “assuming a 10 minute appointment, if a longer appointment the test could take longer” [Q8]<br>“lots more to do in the 10 minute consultation” [Q8]<br>“People can be really upset (angered) by the dementia screening questions. It's worth not rushing.” [Q8]<br>“As a pre-arranged appointment with a single objective” [Q8]<br>“Takes into account prior knowledge of the patient” [Q8]<br>“it depends if you know the patient already & have identified cognitive decline” [Q8]<br>“We are time poor and offering double or review appointments means we are then under more pressure for appts from other patients” [Q8]<br>“If booked as a separate appointment” [Q8]<br>“Needs to be fitted into surgery time” [Q8]<br>“It should be easily accomplished in 1 appt alongside other parts of the consultation i.e. Physical checks” [Q8]<br>“). I find that patients on first presentation may need rapport-building. Often hard of hearing. need time to digest what's going on. However, what usually happens is that there is almost always a social or physical need that presents at the same time. They present because of a problem. “ [Q8]<br>“In the context of a total ten minute appointment” [Q8]<br>“Needs to be 'brief' but for cognitive decline 2 minutes I think should be fine recognising that the brief interventions for smoking for example are different in context.” [Q8]<br>“The test is usually for someone else'sneed/purpose. Diagnosis is a shared interpretive process of weighting up risk and benefit with pt” [Q9]<br>“I don't just use tools to assess is part of assessment” [Q9] | What else needs to happen within the consultation<br>Interpersonal factors<br>Familiarity with the patient<br>Consultation objectives<br>Time allocation<br>Guardian of consultation time<br>Diagnosis as a shared interpretative process<br>GP as the appropriate space for processes |

| | | |
|---|---|---|
| | "it is such a basic scale it can be skewed by anxiety/a bad day etc and it doesn't pick up subtle issues" [Q9]<br>"Most experience with this" [Q9]<br>"Most patients score fairly poorly on it and need referral regardless. However, it seems to do the job for reassuring those with no problems and ensuring onward investigation for those in which there are concerns within the history. Therefore, I find it hard in pitching the outcome at times, to prepare the patient accordingly. Just how trustworthy is the score?" [Q9]<br>"Rarely used an assessment as I have not seen much cognitive decline. Working part-time may contribute to this" [Q9]<br>"Language can be issue. Very intelligent folk score well even when in Decline so feel less confident then" [Q9]<br><br>"They can be useful, but as with depression scores, they're no substitute for a good history. The danger can be that they lead to a tickbox approach" [Q10]<br>"We are often trying to implement these with an aging population, who need time, sufficient hearing and motor function to complete many of them. They often have many issues in the consultation" [Q10]<br>"as a GP I don't believe that I am the best placed person to do it" [Q10] | |
| Generalism vs specialism/expertise | "advice from specialist colleague" [Q5]<br>"done more slowly and accurately by the nurse" [Q8]<br>"I have not seen much cognitive decline. Working part-time may contribute to this." [Q9]<br>"Most experience with this." [Q9]<br>"able to reassure if nromal (and if clincial picture is one of worried well but not if more complex picture /sutle problem or high IQ" [Q9]<br>"I think I can give a reasonable explanation to patient and family" [Q9]<br>"As before but it's a case of "this seems fine we need to think more about your problems (or not)" or "this | General practitioner not specialists<br>Inaccurate in assessment<br>Reassurance role<br>Challenge of complexity<br>Reasonable explanations<br>Non-definitive language<br>Unconvinced by need for screening<br>Lack of knowledge<br>General practitioners not best suited for assessment of dementia<br>Variation in number of patients seen/levels of experience, partly due to part-time working? *Less* |

| | | |
|---|---|---|
| | suggests you would benefit from seeing someone else to review things further." [Q9]<br>". I'm not convinced of the evidence base or practical utility of asymptomatic screening." [Q10]<br>"Would be helpful to have an idea of how to assess premorbid level of functioning for both ends of intellectual spectrum. And how that might affect results of simple tests" [Q10]<br>"I find them a very crude tool and as a GP I don't believe that I am the best placed person to do it." [Q10] | *emergent as an issue than expected*<br>Not a clear link between number of patients seen (up to 30 per month) and dementia speciality in the practice. Surprising |
| Screening | "I don't like the screening aspect" [Q10]<br>"Pressure on time and appointments drives lots of GP decision making Alos increased awareness of memory clinics has driven high demand and increased requests for testing and referral" [Q10]<br>"To clarify, I'm not here talking about screening - that is, where a patient/family/me doesn't have concerns about cognition, but where there is something that had already raised the question. I'm not convinced of the evidence base or practical utility of asymptomatic screening." [Q10] | What does screening mean?<br>Screening as a political drive<br>Increasing pressure on limited resources |
| "doing things properly" | "I'll sometimes arrange a double appointment to do things properly if there's a concern about cognition" [Q8]<br>"nees to be done, and need the best tool available so I would make time" [Q8]<br>"However, what usually happens is that there is almost always a social or physical need that presents at the same time. They present because of a problem. Therefore, for a GP, you need an extra 10-15 minutes on top of the assessment to deal with this. " [Q8]<br>"Able to have longer as can concentrate on assessment only" [Q8]<br>"we have 10 minute appoiintments- these are time consuming consultations, even of brought back for a double appt after bloods it takes | Justification needed to do things properly<br>Professional pride<br>Conflict with the system<br>"what usually happens"<br>Value of 'doing things properly'<br>Hassle/burden of 'doing things properly' |

| | | |
|---|---|---|
| | ages and the informant interview/history" [Q8]<br>"As a pre-arranged appointment with a single objective, to allow a non pressured approach, and to allow me to look out for fluctuating function or attention during the process" [Q8]<br>"It's worth not rushing." [Q8]<br>"done more slowly and accurately by the nurse" [Q8]<br>"they're no substitute for a good history" [Q10]<br>"We are often trying to implement these with an aging population, who need time, sufficient hearing and motor function to complete many of them."[Q10]<br>"I have recently asked care home managers if they would be willing to train staff to do MMSE." [Q10]<br>"I do the longer addenbrookes one for high IQ/young patients but it takes ages and feels like a hassle but can be useful to reassure worried well and avoid a referral " [Q10]<br>"Difficult to assess how much anxiety and depression affects the results" [Q10] | |
| Subtlety | I feel that the ace-iii has good face validity; is better at identifying more subtle presentations off dementia than e.g. MOCA" [Q5]<br>"it is such a basic scale it can be skewed by anxiety/a bad day etc and it doesn't pick up subtle issues" [Q9] | Feel, intuition<br>Basic versus complex scales |
| Reassuring (patients and themselves - GP) | "a normal-range score is more reassuring (as important as its use in identifying actual cognitive impairment)" [Q5]<br>"if only to do the assessment, 15 minutes is ideal (including explanation of what will happen next" [Q8]<br>"I find that patients on first presentation may need rapport-building. Often hard of hearing. need time to digest what's going on." [Q8]<br>"More time to spend in discussion" [Q8]<br>". People can be really upset (angered) by the dementia screening questions. It's worth not rushing." [Q8] | Normal = reassuring<br>Importance of reassuring patients<br>Reassurance *as important* as identifying cognitive impairment<br>Rapport-building<br>Time to discuss<br>Value of time<br>Reassurance and GP confidence in diagnosis<br>Preparing the patient<br>"Worried well"<br>Patient being protected by GP from worry<br>Avoiding referral<br>Also concept of people visiting the GP with concerns; if borderline or |

| | | |
|---|---|---|
| | "Very intelligent folk score well even when in Decline so feel less confident then" [Q9]<br>"it seems to do the job for reassuring those with no problems and ensuring onward investigation for those in which there are concerns within the history. Therefore, I find it hard in pitching the outcome at times, to prepare the patient accordingly. " [Q9]<br>"able to reassure if nromal (and if clincial picture is one of worried well but not if more complex picture /sutle problem or high IQ," [Q9]<br>"As before but it's a case of "this seems fine we need to think more about your problems (or not)" or "this suggests you would benefit from seeing someone else to review things further" [Q9]<br>", I do the longer addenbrookes one for high IQ/young patients but it takes ages and feels like a hassle but can be useful to reassure worried well and avoid a referral" [Q10] | 'worried well' cases, what's driving them to seek help? |
| Relatives | "facilitates relative involvement if required" [Q5]<br>"I think I can give a reasonable explanation to patient and family" [Q9]<br>"I am rarely surprised by results & do it since I am supposed to - discussing problems with patient & carer is more useful to identify dementia" [Q10] | Relatives often missing<br>Relatives to be explained to<br>Patient and carer as a unit<br>Relatives mentioned less than expected |
| Trust | "…I find it hard in pitching the outcome at times, to prepare the patient accordingly. Just how trustworthy is the score?" [Q9] | GP trust of the test score |
| Ideal vs reality | "lots more to do in the 10 minute consultation" [Q8]<br>"It can take 30 minutes though not the 6CIT alone but the whole discussion." [Q8]<br>"As a pre-arranged appointment with a single objective, to allow a non pressured approach, and to allow me to look out for fluctuating function or attention during the process" [Q8]<br>"These tend to be long consultatins without the test and if the patient is reporting memory probs they get fereed irrespective of the result" [Q8] | Tensions between the ideal scenario versus reality<br>Planning in advance<br>**Ideal**<br>Lack of pressure<br>Time to monitor through the appointment<br>Brief test<br>**Reality**<br>Referral decision irrespective of consultation<br>Time pressures – other tests & exams to do |

| | | |
|---|---|---|
| | "it depends if you know the patient already & have identified cognitive decline" [Q8]<br>"again nurse would usually take more time over most things… nurse appts are not resourced " [Q8]<br>"Needs to be quick/ other tests/ exams to do as well" [Q8]<br>"I think a brief assessment tool should be just that, brief. It should be easily accomplished in 1 appt alongside other parts of the consultation i.e. Physical checks" [Q8<br>"Therefore, for a GP, you need an extra 10-15 minutes on top of the assessment to deal with this. Or, if you only have 10 mins, the assessment is the thing that gets squeezed. " [Q8]<br>"other tasks to do" [Q8]<br>"if only to do the assessment, 15 minutes is ideal" [Q8]<br>"less than 10 mins ideally." [Q8]<br>"nees to be done, and need the best tool available so I would make time" [Q8]<br>"Needs to be consistent within the practice, not have different tests used by different staff" [Q8]<br>"it is such a basic scale it can be skewed by anxiety/a bad day etc and it doesn't pick up subtle issues" [Q9]<br>"I don't find this tool particularly discriminating. Most patients score fairly poorly on it and need referral regardless" [Q9]<br>"They need to be short in time to facilitate a realistic assessment in general practice" [Q10]<br>"Stopped using MMSE when charges introduced for using it" [Q10]<br>". I'm not convinced of the evidence base or practical utility of asymptomatic screening" [Q10]<br>"Also MMSE now copyrighted so using out with this" [Q10]<br>"Rebranding on the 'brief' nature of these would be helpful!" [Q10]<br>"if I have enough clinical suspicion, I will refer regardless of high score" [Q10]<br>"The danger can be that they lead to a tickbox approach" [Q10]<br>"More difficult on a home visit" [Q10] | Conflict with other tasks within GP appointment timeframe 'unrealistic'<br><br>Need versus ideal<br><br>Danger, difficulty<br><br>Tick-box approach vs. efficiency/ease of use<br><br>Complexity/detail vs. indiscriminate, easily skewed<br><br>Difficulties in context – home visit vs. in office<br><br>Complexity in mixed presentation with e.g. depression, anxiety |

| | | |
|---|---|---|
| | "Difficult to assess how much anxiety and depression affects the results" [Q10]<br>"discussing problems with patient & carer is more useful to identify dementia" [Q10] | |
| Technology/ process | "Clinical software - in the past it was there as a quick tool until MMSE went" [Q5]<br>"1. I remember its name 2. I can find it on the computer" [Q5]<br>"Available on EMIS system" [Q5]<br>"I have personalised a version of MMSE and embedded it within EMIS so it launches a personalised sheet with the questions pre-printed and then I can answer the score and scan into the record. This makes it easy to administer." [Q5]<br>"if you only have 10 mins, the assessment is the thing that gets squeezed" [Q8]<br>"we have 10 minute appoiintments" [Q8]<br>"I'll sometimes arrange a double appointment to do things properly" [Q8]<br>"again nurse would usually take more time over most things but We pay for nurse appts and time and they are part of the total pressure on the system so nurse appts are not resourced" [Q8]<br>"It's written on the screen... would I remember scoring? Not a chance but I don't need to." [Q9]<br>"We use 6-CIT as per local guidance and it's part of the referral form. It's complex to score but embedded in SystmOne (computer system used by GP, psych, district nurses etc) so quick to score. More difficult on a home visit" [Q10] | Tests available on the computer system (EMIS or SystemOne)<br>Ease of use/access<br>Local guidance, local processes<br>Difficulty outside usual processes (e.g. home visits) |
| Familiarity | With the test<br>"Rarely used an assessment as I have not seen much cognitive decline. Working part-time may contribute to this." [Q9]<br>"I always have to re-google the normal range scores" [Q9]<br>"its the one I know" [Q9]<br>"its the one I use" [Q9]<br>With the patient | Lack of familiarity, experience<br>Part time vs. full time<br>Levels of exposure to cognitive decline<br>The one I know<br>Certainty in practice and/or knowledge<br>Patients' needs<br>Prior knowledge of the patient |

| | | |
|---|---|---|
| | "I find that patients on first presentation may need rapport-building. " [Q8]<br>"it depends if you know the patient already & have identified cognitive decline" [Q8]<br>"Takes into account prior knowledge of the patient" [Q8] | Recognition of cognitive decline through familiarity with the patient |
| Difficult patients – age, education – acknowledged | "I find that patients on first presentation may need rapport-building. Often hard of hearing. need time to digest what's going on. However, what usually happens is that there is almost always a social or physical need that presents at the same time." [Q8]<br>"able to reassure if nromal (and if clincial picture is one of worried well but not if more complex picture /sutle problem or high IQ, Also useful if low score to say - needs referral for further testing- looks like a possible memory issue." [Q9]<br>"More difficult if scores between 25 - 29 but I feel able to use it confidently" [Q9]<br>"it is such a basic scale it can be skewed by anxiety/a bad day etc and it doesn't pick up subtle issues" [Q9]<br>"I don't find this tool particularly discriminating. Most patients score fairly poorly on it and need referral regardless. " [Q9]<br>"Language can be issue. Very intelligent folk score well even when in Decline so feel less confident then" [Q9]<br>"Not always accurate in early dementia or persons with high intellect" [Q10] | The usual patient (presents with multiple needs)<br>The unusual patient<br>Normal levels; worried well<br>Complexity from education levels, high IQ, high intellect<br>Language can be an issue – language used in the test? Or Language used by the patient? Or GP?<br>Patient in decline – a continuum |
| Unwillingness/duress | "Pressure on time and appointments drives lots of GP decision making<br>Alos increased awareness of memory clinics has driven high demand and increased requests for testing and referral" [Q10]<br>"They feel like a huge tiem burden" [Q10]<br>"I do the longer addenbrookes one for high IQ/young patients but it takes ages and feels like a hassle but can be useful to reassure worried well and avoid a referral " [Q10] | Pressure, demand, burden, hassle<br>Increased awareness of memory clinics driving demand<br>Increased numbers requesting tests and referral (to memory clinics)<br>Expertise vs. external pressures<br>Crude tools<br>GP not best placed person – as a role? As an environment? |

| | | |
|---|---|---|
| | "I am rarely surprised by results & do it since I am supposed to" [Q10]<br>"I find them a very crude tool and as a GP I don't believe that I am the best placed person to do it." [Q10] | Is general practice the appropriate place for assessment?<br>"Are GPs using evidence based practice?" In answer to the question this clearly suggests that they are not. GPs not necessarily *able* to follow evidence based practice due to system factors such as System 1/EMS<br>"As a GP I don't believe I am the best placed person to do it" Does this mean they don't think that a GP is in the best position systemically, or that they lack the knowledge, or something else? |
| Language use | "can be useful to reassure worried well" [Q10] | Worried well still in use |

| Glossary | |
|---|---|
| **6-CIT** | 6 Item cognitive impairment test |
| **7MS** | 7 minute screen |
| **10-CS** | 10 point Cognitive Screener |
| **$A\beta_{42}$** | Amyloid beta peptide 42 |
| **ACE (-R/III)** | Addenbrooke's Cognitive Evaluation (- revised/version three) |
| **ADAS-Cog** | The Alzheimer's Disease Assessment Scale - Cognition |
| **ADI** | Alzheimer's disease International |
| **AD** | Alzheimer's disease |
| **AD8** | Alzheimer's disease brief screening tool |
| **AGECAT** | Automated Geriatric Examination for Computer Assisted Taxonomy |
| **AMSTAR (2)** | A MeaSurement Tool to Assess systematic Reviews (2nd edition) |
| **AMT(S)** | Abbreviated Mental Test Score |
| **APEx** | Exeter Collaboration for Academic Primary Care |
| **BCA** | Brief cognitive assessment |
| **BOS** | Bristol Online Surveys |
| **CAMCOG** | The Cambridge Cognitive Examination |
| **CAMDEX** | Cambridge Mental Disorders of the Elderly Examination |
| **CASI** | Cognitive Abilities Screening Instrument |
| **CCCDTD** | Canadian Consensus Conferences on the Diagnosis and Treatment of Dementia |
| **CCG** | Clinical Commissioning Group |
| **CDIG** | Cochrane Dementia and Cognitive Improvement Group |
| **CDPC** | Cognitive Decline Partnership Centre |
| **CDR** | Clinical Dementia Rating |
| **CDT** | clock drawing test |
| **CERAD** | Consortium to Establish a Registry for Alzheimer's Disease |
| **CFAS II** | Cognitive Function and Ageing Study II |
| **CHAIN** | Contact, Help, Advice and Information Network |
| **CHERRIES** | Checklist for Reporting Results on Internet E-Surveys |
| **CLAHRC** | Collaboration for Leadership in Applied Health Research and Care |
| **CPG** | Clinical Practice Guideline |
| **CSF** | Cerebrospinal fluid |
| **DLB** | Dementia with Lewy bodies |
| **DSM-III/ III-R/ IV/ IV-R** | The Diagnostic and Statistical Manual of Mental Disorders (version 3/ version 3 revised/ version 4/ version 4 revised |

| | |
|---|---|
| **EMBASE** | Excerpta Medica dataBASE |
| **FAB** | Frontal Assessment Battery |
| **FAQ** | Functional Activities Questionnaire |
| **FCSRT** | Free and Cued Selective Reminding Test |
| **FTD** | Frontotemporal dementia |
| **GMS-AGECAT** | Geriatric Mental State Schedule - Automated Geriatric Examination for Computer Assisted Taxonomy |
| **GP** | General Practitioner |
| **GPCOG** | The General Practitioner assessment of Cognition |
| **GSA** | Gerontological Society of America |
| **ICD-10/11** | International Classification of Disease – version 10/11 |
| **IQCODE** | Informant Questionnaire on Cognitive Decline in the Elderly |
| **IPA/WHO criteria** | International Psychogeriatric Association/World Health organisation criteria |
| **KICA-Cog** | Kimberley Indigenous Cognitive Assessment |
| **KICA-Screen** | Kimberley Indigenous Cognitive screening tool |
| **MCI** | Mild Cognitive Impairment |
| **MEDLINE** | Medical Literature Analysis and Retrieval System Online |
| **MeSH** | Medical Subject Headings |
| **MIS** | Memory Impairment Screen |
| **MMSE** | Mini mental state examination |
| **MoCA** | Montreal Cognitive Assessment |
| **MRI** | Magnetic Resonance Imaging |
| **NHS** | National Health Service (UK) |
| **NIA-AA** | National Institute for Aging and the Alzheimer’s Association |
| **NICE** | National Institute for Health and Care Excellence (UK) |
| **NIHR** | National Institute for Health Research |
| **NINDCS-ADRDA** | National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association |
| **NINCDS-AIREN** | National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l'Enseignement en Neurosciences |
| **NINCDR-CERAD** | National Institute of Neurological Disorders - Clinical Dementia Rating -Consortium to Establish a Registry for Alzheimer’s Disease |
| **PAS** | Psychogeriatric Assessment Scale |
| **PBKDF2** | A password hashing algorithm |
| **PCL** | Prueba cognitive de leganes |
| **PROSPERO** | International Prospective Register of Systematic Reviews |

| | |
|---|---|
| **PsychInfo** | Database of abstracts of literature in the field of psychology. |
| **ROBIS** | Risk of Bias in Systematic reviews |
| **RUDAS** | The Rowland Universal Dementia Assessment Scale |
| **SASSI** | Short and Sweet Screening Instrument |
| **SHA256** | A cryptographic hash function |
| **SIS** | Six item screener |
| **SPMSQ** | Short portable mental status questionnaire |
| **SSL** | Secure Sockets Layer |
| **TICS** | Telephone Interview for Cognitive Status |
| **TRIP** | Turning Research Into Practice database |
| **TYM** | Test Your Memory |
| **UK** | United Kingdom |
| **VAD** | Vascular Alzheimer's Disease |
| **VF-an** | Verbal Fluency - animals |
| **WHO** | World Health Organisation |

# 7. Discussion

*"On veut le beurre, l'argent du beurre et baiser la crémière"*

*[We want the butter, the money for the butter, and to kiss the milkmaid]*
*- French proverb*

There is no simple answer to the question "how can the accuracy of BCAs be improved when used as part of the process for identifying dementia in general practice?". The lack of clear evidence on the most accurate BCAs for use in primary care is an inconvenient truth, as much of the discussion around improving diagnosis rates presupposes that the best tool for diagnosing dementia in general practice is well established. This is not the case. There are, however, a range of measures that could contribute to improving the accuracy of these assessments within primary care. By scrutinising current evidence, generating new survey data on how GPs use BCAs, and critically analysing research methodology, this research contributes to how we might better understand and improve the accuracy of BCAs when used as part of the process for identifying dementia in general practice.

This chapter discusses the findings made throughout this process, what these findings mean in practical terms, and makes recommendations for research and practice. The chapter concludes with reflections on the PhD as a whole, considering what has been learned and acknowledging those who have helped along the way.

## 7.1. Aim and objectives

The primary aim of this thesis was to address the question "how can the accuracy of BCAs be improved when used as part of the process for identifying dementia in general practice?" by establishing a clear picture of the evidence for diagnostic accuracy of BCAs for identifying dementia within general practice.

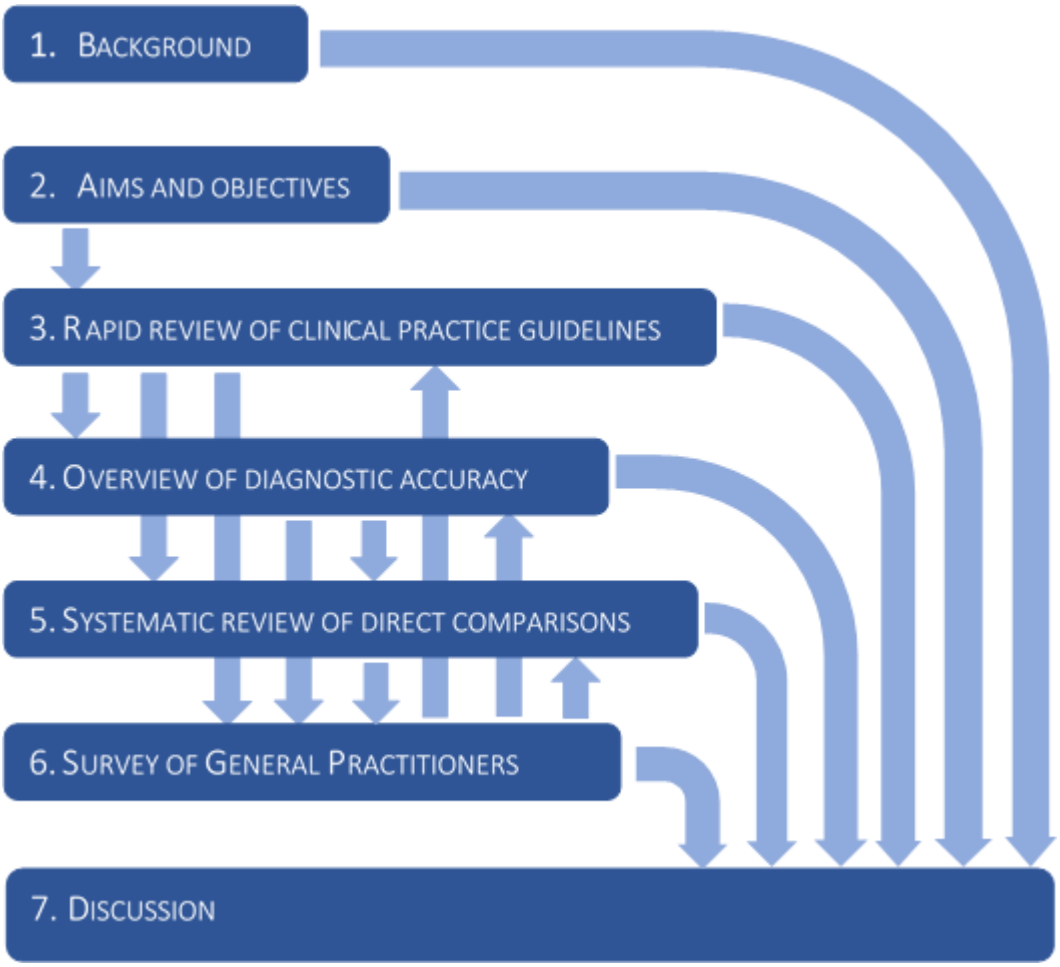
The objectives of this thesis were:

- To assess the diagnostic accuracy of BCAs used to identify dementia specifically in primary care;

- To review the diagnostic test accuracy evidence by assessing studies that have directly compared GPCOG and MMSE for identifying dementia in general practice;
- To explore the views of general practitioners around dementia screening tests; and
- To make practical recommendations for improving the accuracy of BCAs, specifically for use in general practice based on evaluation of the evidence in the objectives listed above.

The phases of this work across PhD chapters and how they interact are shown in Figure 31.

Figure 31 PhD thesis by chapter: how results from chapters of the thesis informed other chapters



## 7.2. Chapter summaries

### Chapter 1: Background

Dementia is under-recognised globally, and there are many gaps in understanding how to improve dementia diagnosis in general practice. Many national strategies promote earlier or more timely diagnosis of dementia within primary care. Paradoxically, recommendations for the best tools to use as part of this process are unclear, vary widely and – where data are presented – the evidence base is mixed. BCAs are the main tool available to family doctors for measuring cognitive performance, but evidence of their diagnostic accuracy is unclear. It is essential to understand the diagnostic accuracy of BCAs when used as part of the diagnostic process for identifying dementia in general practice in order to judge their suitability and make clear clinical recommendations on their use. As a wider aim, greater understanding of the challenges of identifying dementia in general practice and – in particular – of the assessment of cognitive function as part of a wider diagnostic assessment for possible dementia is needed as a priority.

### Chapter 2: Aims and objectives

The primary aim of this thesis was to address the question “how can we improve the accuracy of brief cognitive assessments when used as part of the process for identifying dementia in general practice?” by establishing a clear picture of the evidence for diagnostic accuracy of brief cognitive assessments for identifying dementia within general practice.

This was addressed through the following objectives:

- establishing current clinical practice guidelines relating to the use of brief cognitive assessments used to identify dementia in primary care;
- reviewing the evidence on the diagnostic accuracy of brief cognitive assessments used to identify dementia in primary care;
- assessing the diagnostic accuracy evidence in studies that have directly compared GPCOG and MMSE for identifying dementia in general practice; and
- exploring the views of GPs around BCAs for identifying dementia in general practice ,and triangulation of survey data with other identified evidence.

This assessment of the evidence allowed the building of practical recommendations for improving the accuracy of brief cognitive assessments specifically for use in general practice.

### Chapter 3: Rapid review of CPGs relating to the use of BCAs used as part of the process for diagnosing dementia within general practice

This rapid review was invaluable in identifying what CPGs published in the English language are available for advising BCA use in general practice. The review also established a high degree of variation across the three CPGs identified, and found a paucity of English language CPGs from other countries. The lack of consistent recommendations amongst BCAs in terms of the tools themselves, in their application and in their selection was an important discovery and worth further investigation. In addition, the scarcity of guidance given within the identified CPGs on tailoring BCA choice and use for specific populations was particularly notable given the three countries where CPGs were identified (Australia, Canada and the UK) all have cultural and linguistically diverse populations where tailoring of tools would be highly appropriate. This rapid review indicates that greater clarity and consistency is needed from CPGs relating specifically to the use of BCAs as part of the process for identifying dementia in general practice.

### Chapter 4: Overview of diagnostic accuracy

The systematic review evidence identified within the overview on the accuracy of BCAs was overwhelmingly focussed towards assessing the MMSE despite diminishing clinical relevance and reduced recommendations for use in general practice. Highly selective study populations did not relate consistently to general practice populations, even where ‘general practice’ was the stated population of interest. The prevalence of dementia in included study populations ranged from 4% to 51%. BCA thresholds varied widely, and were not clearly reported in all cases. Standardised diagnostic accuracy data (sensitivity and specificity) were not explicitly reported in all reviews, and in many cases was pooled across included studies without disaggregated information provided. Variability is amplified in overviews. Inconsistency across populations, settings, tests and thresholds made comparison across different reviews very challenging. In several cases such as those illustrated above, there was insufficient evidence found to draw conclusions on the accuracy of

particular tests for identifying dementia in primary care, and this lack of evidence needs addressing as a priority.

### Chapter 5: Systematic review of direct comparisons

Direct within-study comparisons of two BCAs (MMSE and GPCOG) were possible but highly limited due to variation in sample size, study design, language and different MMSE thresholds used. Based on the limited evidence of mixed quality studies in the review, MMSE and GPCOG Total perform similarly across both sensitivity and specificity when used to identify patients with possible dementia in a general practice setting. The review highlighted the need for better conducted studies.

### Chapter 6: Survey of General Practitioners

The survey of 52 UK general practitioners was designed to reflect the reality of general practice experiences, in order to contrast this with findings generated by evidence syntheses and review of the literature. Eighty-nine per cent of GPs surveyed did not have access to dementia specialists in their practice. The most commonly used BCAs amongst surveyed GPs were the MMSE and GPCOG (both at 32%), 6-CIT (17%), and the Clock Drawing Test (7%) with Mini-Cog, MoCA, ACE-III and IQCODE all chosen by one respondent each. GPs reported confidence in interpreting tests results and explaining them to patients. A number of potential barriers to clinical practice identified by respondents were the length of time a test takes to administer, ease of use and the perceived appropriateness of screening in general practice. Triangulation of the survey findings with identified evidence elsewhere in the thesis and evidence already published in the literature led to several insights: CPG use was only mentioned in 6% of GPs' survey responses when asked what affected their choice of BCA, and this was in line with the wider literature where CPGs did not feature regularly in features affecting GP choice of measures. A key finding of the survey that resonated strongly with the results of the systematic review and overview were considerations of associated factors beyond the accuracy of tests, such as the time the test took to administrate, patient understanding and further discussion of prognosis and management options. These results were strongly supported within the literature around influential factors beyond accuracy that are an added concern for GPs when conducting an assessment for dementia.

In summary, whilst a heightened need to identify dementia in primary care is recognised in policy and research, the research presented in this thesis shows:

- Clinical practice guidelines are incomplete, inconsistent and uncommon;
- Despite a wealth of research, there is still no clear evidence to recommend any specific BCA for use in primary care;
- Whilst evidence is still lacking on the accuracy of the most common BCAs being used in primary care, accuracy itself is not all that counts; administration time and confidence in using the test are important;
- GPs recognise these issues alongside the need for greater understanding of patient factors (multi-morbidity, education levels and treating patients with particular needs, such as physical impairments or where English is not their first language) and improved management processes such as the integration of tests into practice software;
- Many GPs also question the role of screening and case finding in general practice.

This is the first rapid review of CPGs directly focussed on the use of BCAs to be used as part of the process for identifying dementia in general practice.

As far as the author is aware, this is also the first systematic review to assess direct comparisons of the diagnostic accuracy of MMSE and GPCOG used as part of the process to identify dementia in primary care.

This is also the first overview looking at BCAs for identifying dementia in primary care, and through this several issues have been identified which contribute to the development of overview methodology.

## 7.3. Comparison with what is currently known

In comparison with the wider literature, these findings are broadly in line with current evidence with some notable exceptions.

### 7.3.1. BCAs and Clinical Practice Guidelines

The rapid review identified a number of clear recommendations for practitioners, whilst recognising the limitation that only CPGs in the English language were identified. Taking this into account, there was still a lack of diversity and breadth found in CPGs published in the English language. The three countries from which CPGs were identified (Australia, Canada and the UK) all have culturally and linguistically diverse

populations, and whilst the Australian CPGs recognised BCAs tailored for particular groups there was no mention of tailoring or different population needs within the BCAS for Canada or the UK.

Greater agreement between CPGs will help improve adoption, clinical practice and understanding of the place of BCAs within the diagnostic assessment in a non-specialist setting. Whilst this agreement is needed, it needs to be seen in the context of understanding which BCAs are best for these requirements, and the place of BCAs within general practice. To improve consistency and standardisation in the use of CPGs these questions first need to be answered.

The majority of existing BCAs demonstrate insufficient validation and replication of for general practice use. In order to assess BCA suitability for different general practice populations, variations in setting, and performance in identifying differing levels of severity of cognitive impairment these validation and replication studies need to be completed.

Once there is greater understanding of the critical features of BCAs for general practice use, it is highly likely that CPGs will be far more available, coherent and cohesive for the intended general practice audience.

The most recent dementia clinical practice guidelines issued by NICE in the UK[383] support the use of the

- 10-point cognitive screener (10-CS),
- 6-item cognitive impairment test (6CIT),
- 6-item screener,
- the Memory Impairment Screen (MIS),
- the Mini-Cog and
- Test Your Memory (TYM)

as part of the process for identifying dementia in primary care. Members of the Guideline Committee stated that these tests were broadly similar, and did not recommend one over another[384].

In contrast to the NICE guidelines, study-level evidence was found supporting the use of Mini-Cog for people where English was not their first language, and for people with

low levels of education[144] but no clear systematic review-level evidence to support the use of Mini-Cog as a tool for use in primary care.

This finding is in line with conclusions of the authors of the single primary study[255] cited within the NICE CPG[383] which advised "against the use of the Mini-Cog, particularly in settings where the proportion of unrecognized dementia is low"[255]. There was no evidence presented within the NICE guidelines on the use of GPCOG, whereas this was found to be the second most popular test used by GPs in the survey, and the second most commonly used within research literature for the evidence syntheses.

These NICE guidelines vary from the tools recommended by the Gerontological Society of America (GSA) in 2017[66,67] – the Mini-Cog, GPCOG, MIS, Short IQCODE and AD8.

Within the overview summarising the accuracy of BCAs, no systematic review-level evidence was found on the diagnostic accuracy of 10-CS, 6CIT, the 6-item screener, or TYM for identifying dementia in primary care. In support of recommendations made, NICE guidelines cite: one study for 10-CS[385] based in secondary care; one study for 6CIT[386] based in secondary care; one study for the 6 item screener[387] based in secondary care; and two studies for the Test Your Memory[388,389] both based in secondary care. Alongside this evidence, the NICE Guideline Committee based these recommendations on their 'experience and opinion'[384].

In line with the 2017 survey of 52 GPs in the UK finding MMSE, GPCOG and CDT as the most popular BCAs reported in this thesis, a contemporary survey of 445 doctors across 25 European countries[390] found the MMSE and CDT were the most common BCAs used by primary care practitioners. This Europe-wide survey also found that GPCOG was popular amongst GPs in the UK, and MMSE was the recommended BCA in many countries; in 12 out of 25 countries it was mandatory to conduct the MMSE before prescription drugs for dementia could be prescribed. The UK based survey findings reported in this thesis broadly agree with this wider European survey, as well as findings from earlier surveys based in the USA[391] and Canada[339].

A recently published systematic review by Chan and colleagues[392] compared the diagnostic accuracy of computerised BCAs against pen and paper BCAs, with the primary aim of MCI diagnosis and the secondary aim of dementia diagnosis. Participants were recruited from a number of different populations and results were

pooled, making disaggregation impossible without contacting authors for original data. However, as respondents within the GP survey indicated a preference for computerised tests, the distinction between manual and computerised tests would be worth exploring in terms of administration time and acceptability for a general practice population.

One factor not assessed within this thesis was the potential value of using BCAs in combination with other tests such as CSF biomarkers ($A\beta_{42}$, total tau, and p-tau) and imaging biomarkers measured via MRI scans. A recent multicentre study combined these factors for identifying dementia compared to no dementia, and for identifying subcategories of dementia such as Alzheimer's disease (AD), frontotemporal dementia (FTD), vascular dementia (VAD) and dementia with Lewy bodies (DLB)[393]. The study authors found that included cognitive tests (MMSE, memory (learning), memory (recall), Trail Making Test A, Trail Making Test B, Animal fluency) showed fair or good performance in separating people with dementia from controls when used in isolation. Cognitive tests were unsurprisingly found to be poor at discriminating between subtypes, probably due to insufficient discriminatory power and symptomatic overlap across some subtypes. Improved diagnostic value was reported when combining BCAs, CSF biomarkers and MRI to differentiate between AD and VAD, AD and DLB and FTD and DLB – although as the authors testify these results provide more of an indication about the potential additional value of combining measures, rather than definitive guidance on optimal combinations. Understanding of the diagnostic significance of biomarkers for dementia and subtypes is still developing, and the additional value of biomarkers as part of the diagnostic process for people with possible cognitive impairment is not yet proven[394].

### 7.3.2. Availability and preferences around BCAs

In much of the discourse around dementia diagnosis, there is the implicit assumption that improving levels of investigation for possible dementia within general practice will naturally lead to increased diagnosis rates. The evidence does not yet support this assumption, and may increase the degree of misclassification.

A recent UK trial[92] assessed the effectiveness of an intervention designed to help improve diagnosis rates within a general practice population using letters from GPs sent to their patients aged 70 or over without a diagnosis of dementia inviting them in for assessment. The intervention prompted a significant increase in the proportion of

patients consulting with their GPs over suspected memory problems, compared to patients in the 'usual practice' control arm (odds ratio = 1.4; 95% Confidence Interval = 1.28, 1.54) but no increase in referral on to memory clinics for diagnosis. The study did not show whether participants were without symptoms, were symptomatic with other causes, or whether GPs were not able to correctly identify cognitive impairments or dementia. Future research could usefully unpick these potential underlying reasons for increased consultations not translating to increased referrals or diagnoses.

In contrast to these findings, another study[395] conducted in the wake of NICE 2006 guidance on dementia diagnosis[396] and the 2009 UK National Dementia Strategy[397] explored whether there was an increase in the use of cognitive assessments reported in referrals from primary to secondary care over the period encompassing the launch of these guidelines, compared to use of assessments before guidelines were issued. The study authors found that, over a two year period, the number of GP referrals rose but the percentage of dementia diagnoses fell. The rate of cognitive assessment use did not change over the two years. The authors suggest this is due to guideline pressures driving up the number of healthy people visiting their GP, resulting in no new (or fewer) diagnoses. There are other potential reasons, such as a lessening of practitioner confidence in available tests and testing (resulting in fewer BCAs being conducted), a lack of clarity around dementia treatment and management options (meaning clinicians are less likely to initiate tests or further testing and patients are less likely to follow GP advice or seek follow-up appointments), or symptomatic presentation becoming more complex (for example, comorbidities making diagnosis a more complicated process). There may be many other explanations or a combination of some already posed, and this is another area where future research would benefit clinical practice by exploring the reasons underlying observed increases in presentation and decreases in diagnosis.

Several studies have highlighted the difficulty of using clinical diagnosis as a reference standard within research, where dementia diagnosis is complex, shows substantial comorbidity with other conditions[398], and agreement between clinical diagnosis of dementia and diagnosis using post mortem pathology is between 70% and 90%[399,400] – leaving between 10% and 30% of diagnoses in doubt. Comorbidities are a real challenge to diagnosis, with one study reporting a large minority of people with dementia have other unaccounted for underlying pathologies (20% to 40%)[401].

Understanding common comorbidities and how they may affect clinical presentation when investigating possible dementia would improve understanding of dementia diagnosis.

A thought-provoking new finding reported in an abstract from the UK-based Cognitive Function and Ageing Study II (CFAS II) Dementia Diagnosis Study[402] suggests there may be little demonstrable improvement in health outcomes for people *with* a diagnosis of dementia compared to those *without* a diagnosis of dementia. The reasons behind this may not be obvious, in that it may be an artefact of measurement (i.e. the improvement of health outcomes are not captured within the study), design, or some other aspect not directly related to a lack of actual benefit to the individual. The possibility remains, however, that there is little measurable benefit for health outcomes, and this should be explored as a priority in order to improve and progress the wider discussion of dementia diagnosis, where improved health outcomes following a diagnosis are often assumed.

### 7.3.3. Findings from GP surveys

Continuing the theme of assumed benefit versus existing evidence, the concept of timely diagnosis is now well-established within mainstream dementia discourse[7,22,26,27,46,48,73,332,403,404]. 'Timeliness' or a timely diagnosis has several nuanced definitions based upon either a readiness to undergo assessment directed by the person *and* their family[15,73,314,404] or when the patient wants it *or* when the carers need it[7,27,348,405]. The 2009 Nuffield Council on Bioethics paper 'Dementia: ethical issues'[78] makes it clear that timeliness should be dictated by the patient and that "diagnosis is likely to be timely at the point when the cognitive and other changes they are experiencing begin to affect their lives and the lives of people close to them". The report also cautions that timeliness should be dictated by the person with suspected cognitive issues and their family[78] rather than any outside agency.

The GP survey found respondents had concerns around the timeliness of the diagnosis, alongside questioning whether general practice was the most appropriate place for dementia diagnosis. The importance of a timely diagnosis is heavily emphasised across national strategies and international guidelines, including recent reports by the World Health Organisation[406] and Alzheimer's Disease International[26], yet the evidence of benefit such as improved health outcomes, as discussed above[402], is less well established.

An often-stated benefit of timely diagnosis is the empowerment of patients to plan for the future. In a recent study from Israel which interviewed people with dementia and their families to explore their perceptions of the barriers to timely diagnosis, and benefits of diagnostic evaluation[407], patients stated a number of perceived benefits such as starting medication, reducing anxiety, confirming suspicions and increasing awareness of possible treatments. None of the 26 patients and 27 family members mentioned planning for the future as a benefit, and some of those interviewed found no benefit in diagnosis at all.

This patient perspective is often overlooked, and was not assessed within the survey. A recent Australian study[15] which conducted in-depth interviews with nine people with dementia found a great deal of variation in views and preferences for diagnosis and the processes surrounding diagnosis. Diversity amongst people with suspected dementia is rarely acknowledged amongst guidance and research, yet emerged within this study as one of the most important factors to recognise when planning services. The important and often under-recognised role of GPs in helping people negotiate those services was highlighted within the abovementioned Australian study, as well as another study of interviews with people with dementia in the UK[380]. This finding was echoed by respondents in the GP survey.

The lack of definitive BCAs appropriate for use in general practice, and limited access to diagnostic tools, are barriers to diagnosis reported within the wider literature[19,48,343,408,409] and echoed within the GP survey and evidence syntheses. Also in common with respondents from the survey, many GPs report that they often favour clinical judgment over formal tests when assessing someone for dementia[410-412].

The potential benefits of case finding are unclear to many GPs, and the distinction between case finding (as supported by the Alzheimer's Society and UK NICE guidelines[61,109]) and screening has come under intense scrutiny in recent years[31,85,413-416]. One recent study that sought to assess the potential benefit of screening all patients over the age of 75 for dementia beyond those patients already identified by 'passive case-finding' found that symptomatic case finding was better than screening[81], but these findings were limited by an imperfect reference standard (CAMCOG) and an artificial testing scenario which did not translate to a standard GP consultation[81].

No benefit was found in a recent RCT exploring the effect of an educational intervention targeting case-finding and subsequent care on diagnostic yield and patient mental health[417], and the authors recommended against the use of case finding within family practice but instead suggested improving diagnostic procedure for symptomatic patients. In another study exploring the effect of case-finding in hospitals with primary and secondary care practices, GPs reported barriers to practice in terms of lack of access to hospital data, and lack of clarity in roles and expectations[418], and these concerns were echoed within the survey of GPs with a number of respondents voicing concern over the validity of case-finding policies. The UK National Screening Committee[83] and US Preventative Services Task Force[419] have made clear policy statements that do not support the use of screening for dementia within primary care. The UK National Screening Committee is currently reviewing guidance on dementia screening, and latest published evidence is supportive of current recommendations[420]. From limited research to date on the impact of policies aiming to increase dementia diagnosis rates in general practice, there is insufficient evidence of effect to support further use of case-finding in primary care[417,421].

Lack of time was found to be a barrier to formal diagnosis in much of the evidence[30,390], in line with responses to the GP survey. Practitioners want more time for consultations, and a shorter BCA test. Partly attributed to the lack of time available, GPs report often relying on personal judgment to assess patients for potential dementia[324,411].

## 7.4. Limitations

### 7.4.1. Limitations relating to BCAs

Within this PhD project, the specific focus was on BCAs for use as part of the process for identifying dementia in primary care and general practice. The definition of “BCA” has proved difficult and contentious, as discussed within Chapter 3. There is no consensus definition within the research or clinical literature. The use of the word “brief” is commonplace and often used to refer to tests taking ten minutes or fewer to administer[88,116,225,303,395], but whilst test administration time is often stated, empirical measurement is rarely reported [119,422]. This lack of evidence to support common practice means the term ‘brief’ can be very misleading, whatever the intention in its’ use.

This emphasis required judgment on what constituted a BCA and what was suitable for primary care and general practice, as well as exclusion of other conditions such as

Mild Cognitive Impairment. a conscious decision was made to comply with a generally-accepted concept of BCAs as 'dementia tests' or tests for identifying dementia, rather than the more accurate specification of BCAs as cognitive assessment tools, with capacity to assess particular aspects of an individual's cognitive function. Dementia as a syndrome is far more complex than a discrete disorder of cognition[423], yet all the tools assessed within this thesis are primarily tools that assess cognition in various ways. Some, such as those with informant components like GPCOG, gather additional information on functional abilities - but the underlying premise of this PhD thesis is that BCAs are conceptually-valid tools to be used as part of the process for identifying dementia. It is recognised that this is not an irrefutable position, and there is increasing discussion of the value of taking a more pragmatic approach to dementia identification[424], where the abilities and disabilities of the individual are most prominent and prioritised beyond making diagnoses – at least until a time that the label is indisputably useful to the individual and their families[425-427].

The cost-effectiveness of using different BCAs as part of the process for identifying dementia in general practice was not explored as part of this PhD thesis. This would provide valuable information for commissioners as well as potentially useful additional data for practitioners when deciding between different BCAs, but was beyond the scope of this thesis and would warrant a separate study by itself. Recent work looking at the cost-effectiveness of using GPCOG, MMSE and 6-CIT for detecting dementia in primary care[428] found that GPCOG provided the most cost-effective option compared to clinical (GP) judgment alone, due to earlier access to medications. It would be valuable to unpick some of the assumptions made within this work, for example that all diagnoses reached using these BCAs would be 100% accurate – which we know is not the case. The complexities of dementia as a syndrome and different dementia subtypes were not accounted for within this research, and different downstream treatment options beyond drug therapy were also unaccounted for within the model. Still, this work demonstrates a useful starting point for future investigations of cost effectiveness of the process for identifying dementia within general practice.

It may have been valuable and provided more population-specific data if the condition of interest had been extended to include Mild Cognitive Impairment (MCI) as well as dementia. Whilst MCI is sometimes referred to as 'pre-dementia'[429,430], conversion from MCI to dementia is not guaranteed[431,432]. Some studies show a 60% conversion

rate from MCI to AD within a population-based sample[433] but there is little current evidence of conversion rates within a general practice sample. It was therefore concluded that the link between dementia and MCI is insufficiently established or understood to include MCI as a condition of interest alongside dementia within the evidence syntheses.

### 7.4.2. Limitations relating to research methods

*Rapid review*

The rapid review of CPGs was conducted as a highly specific and targeted evidence synthesis, to be conducted within time and resource allowances to enrich the understanding of current CPGs for this thesis. It addressed the research question fully and was highly valuable for greater comprehension of current BCA guidance, yet given more capacity for search specialisms and additional reviewer resource this could have covered the question to a greater depth and breadth, for example removing the English language restriction and translating texts identified in languages other than English.

*Settings and prevalence*

The prevalence of a condition or disease is the presence of that disease in a particular population at a certain point in time. This is often contrasted with incidence, which is the number of new cases of a condition or disease in a particular population within a certain time period86.

Many of the most established and timeworn cognitive tests were first validated in a clinical population such as a memory clinic where the prevalence of dementia is far higher (up to 56%[434]) than in a normal primary care population where the prevalence is around 7%[90 70,113,435]. This means that in a memory clinic population (i.e. people registered with a specialist memory clinic) out of every 100 people 56 will have dementia – compared to a prevalence of 7% in general practice population where for every 100 people within a general practice area, 7 people will have dementia.

Generally people attending a memory clinic will be older adults, and will have memory-related symptoms which brought them to the clinic in the first place. In contrast a primary care population is a more mixed age-group (16.9% over 65 years old, 7.8% over 75 years old and 2.2% over 85 years old for the England General Practice Average[435]) presenting with a variety of symptoms to their GP.

This difference in prevalence matters, because it has a substantial impact on the accuracy of the test being used. Below is a worked example to illustrate this point.

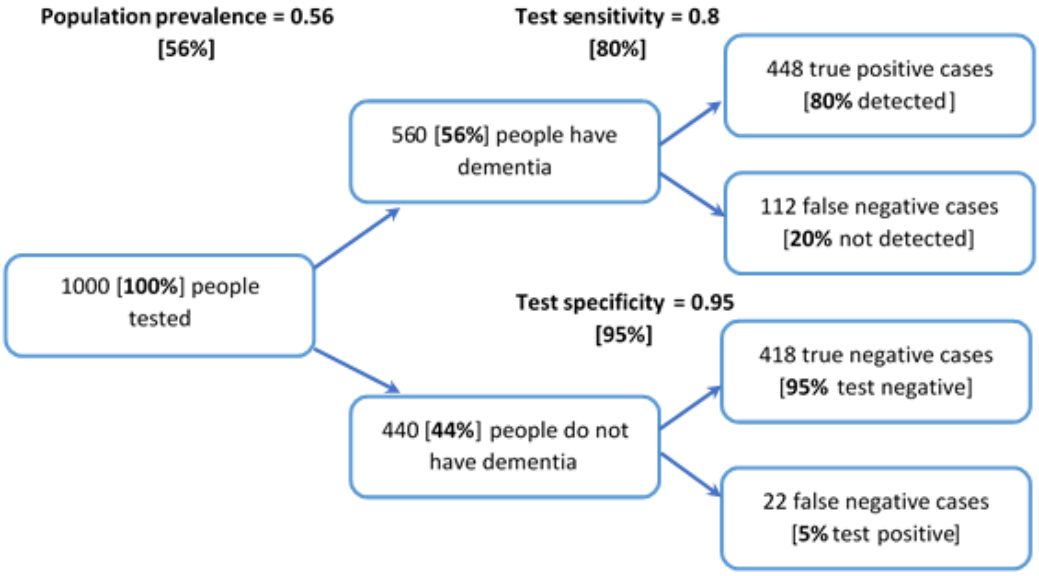
Imagine a test for dementia with 95% specificity (i.e. 95% of people without dementia will be correctly diagnosed as not having it, specificity = 0.95) and 80% sensitivity (i.e. 80% of people with dementia will be correctly identified as having the condition, sensitivity = 0.80). These sensitivity and specificity scores are better than many tests currently used in the health care system and sound fairly good to many people.

Imagine that we test 1000 for dementia from a memory clinic population with a dementia prevalence of 56% as shown in 32.

If we test 1000 people from a memory clinic population with a prevalence of 56%, 560 (56%) will have dementia and 440 (44%) will not. Of the 440 people without dementia, 418 (95%) will correctly receive a negative test result (true negative) whereas 22 people (5%) will receive an incorrect positive test result, i.e. they will be told they have dementia when they do not (false positive). Of the 560 people with dementia, 448 (80%) will be detected and 112 (20%) will be missed.

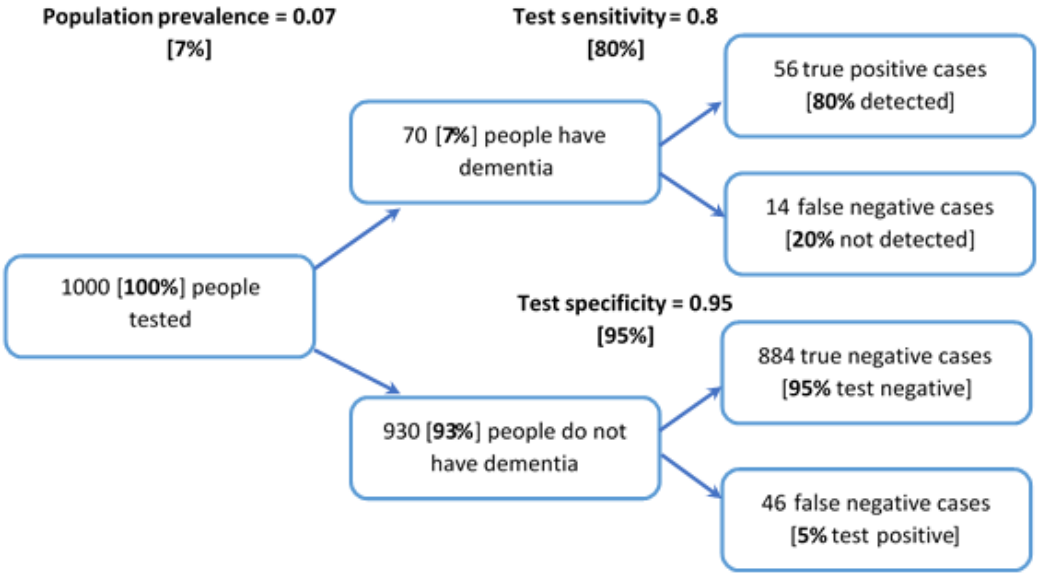
In combination, 448+22= 470 people will test positive for dementia, of which are 22 false positives. Therefore the false discovery rate is 22/470= 4%. This means that if you test positive for dementia within this population, the chance that you really do have dementia is 96%, which is a very high discovery rate.

Figure 32 Tree diagram to illustrate the false discovery rate in diagnostic tests used in a memory clinic population




Now imagine that we test 1000 people for dementia from a general practice population with a prevalence of 7%. As shown in Figure 33, 70 (7%) will have dementia and 930 (93%) will not. Of the 930 people without dementia, 884 (95%) will correctly receive a negative test result (true negative) whereas 46 (5%) will receive an incorrect positive test result, i.e. they will be told they have dementia when they do not (false positive). Of the 70 people with dementia, 56 (80%) will be detected and 14 (20%) will be missed.

Figure 33. Tree diagram to illustrate the false discovery rate in diagnostic tests used in a general practice population




In combination, 56+46 = 102 people will test positive for dementia, of which 46 are false positives. The false discovery rate is 46/102 = 45%. This means that if you test

positive for dementia within this population, the chance that you really do have dementia is only 55% - slightly better than chance.

*Population*

From the illustration above, it is clearly important to take into account the population within which the test is intended to be used when choosing between tests, yet guidance is often unclear about the suitability of different tests for different population[159,436].

Within the overview, there was a lack of consistency amongst the reporting of both individual study populations and systematic review populations. The population of interest as stated in the overview protocol was “primary care” and “general practice”, and this was specified in search terms and inclusion criteria at the title and abstract screening stage. When study data were investigated, they were clearly from highly selected clinical populations with reported disease prevalence within study samples as high as 51%[214], This figure is far more suited to the dementia prevalence of a highly selected memory clinic population compared to a dementia prevalence of around 6.5% in general practice[437]. With this level of variation and inconsistency within the population data, it offers a poor opportunity to fairly compare across review evidence.

In the systematic review, studies excluded participants on the basis of common diagnoses such as delerium[142,288], visual, hearing and physical impairments[142,288,438] and depression[142]. These are comorbidities regularly found within a general practice population and in people living with a diagnosis of dementia[439-441], and are present amongst the complexity of conditions and confounders when assessing cognitive impairment in this population[442,443]. Excluding these conditions therefore risks destabilising the generalisability and applicability of findings to general practice.

*Identification of evidence*

Searches were developed for the overview by identifying key terms within background scoping papers[29,42,45,130,379,391], consulting information specialists and running exploratory searches using different terms. Whilst this was a logical approach, studies may have been missed which used tests that did not make explicit reference to cognition in the article title or abstract. Equally studies may have been overlooked that were conducted with family practitioners, as searches were only conducted for primary care and general practice, as well as ‘community’ within the Cochrane Library. There

is the potential that papers were missed due to the searches employed, but no papers have yet been identified that were missed via other components of the search strategy such as forward or backwards citation chasing that would have been suitable for addressing the aims of the thesis. As a result it may be comfortably concluded that the searches were well designed.

The overview methods used to explore systematic review data summarising the accuracy of BCAs for identifying dementia in primary care were exploratory in some areas, and without clear precedent. There were areas where new and emerging methodological approaches could have been explored more thoroughly by contacting more authors in the field of overview methods development in order to make methodological decisions based upon the latest advances. For example, assessing study quality and risk of bias was not well established within overviews of diagnostic accuracy, and using ROBIS and AMSTAR meant there was some overlap in process but also gaps in assessment such as the inability to tailor assessments and the AMSTAR question 10 "Was the likelihood of publication bias assessed?"[444]. This specific question carries a negative weight for a negative answer, i.e. if the likelihood of publication bias was not assessed, the quality of the review is marked down. This is not in line with current guidance from the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[75], which advises that further understanding of the interaction between publication bias and asymmetry within diagnostic accuracy research is needed before publication bias assessments can be applied in the same way as in systematic reviews of interventions[445,446]. These methods have been reflected upon with colleagues from the Cochrane Collaboration[275,447], and we have highlighted a number of methodological decisions, alongside key features and objectives, which may be of benefit to others for considering during the development of an overview protocol. A recent update to AMSTAR, AMSTAR2[448] has substantially reworked a number of the original questions in order to encompass randomised and non-randomised study evidence – but this tool is still explicitly not designed for diagnostic accuracy data. A new piece of research is underway assessing the usability, applicability and reliability of the AMSTAR, AMSTAR2 and ROBIS tools[449] and this may provide valuable insights into how these tools compare. If this exercise can be related specifically to diagnostic accuracy data this would be a useful addition to current knowledge.

In a 2011 systematic review of the accuracy of family physicians' diagnoses of dementia[450], similar issues were found to those reported regarding imperfect reference standards, unaccounted missing data, unclear timings between administration of the index test and reference standard, lack of blinding in interpretation of tests, and non-representative patient populations. In the context of a systematic review, it would have been possible to pre-state exclusion criteria based upon some of these issues alongside specifying studies would only be included if they used a particular reference standard reported with a base set of information. This may well have resulted in a higher quality but a smaller pool of study-level information, and in the context of the overview, the available systematic review-level evidence would potentially be even more limited. It is unlikely that this route would have resulted in identifying data sufficient for meta-analysis, although this may have provided better quality information which would have allowed firmer conclusions to be drawn, allowing for more directive practice recommendations.

*Discordant results*

It was not possible to explore, even partially, the reasons behind discordant results due to the lack of data. For example, if Mrs Smith has a positive test result on the MMSE, but a negative test result on the GPCOG, why is this? What is the difference (or differences) between the tests, and what are they actually identifying? These factors could probably be investigated from existing datasets using personal identifiers for individual participants, and it would be of great value to follow the testing journey of specific people in order to tease out potential intra-test variations, and explore some of these greater complexities in the testing pathway. If this could be combined with knowledge of other test factors then the spectrum of understanding around the testing process would significantly broaden.

Diagnostic accuracy is a narrow indication of test performance[451] with outcomes influenced by key aspects such as disease prevalence [452], yet is frequently measured and (sometimes poorly) reported[452-456]. Quality issues are common around technical aspects of diagnostic accuracy such as the reporting of accuracy estimates, but also more simply in reporting of inclusion criteria and sampling details[455]. Context is often missing in the reporting of diagnostic accuracy studies, yet is a fundamental aspect of understanding how diagnostic accuracy fits within the clinical pathway[457]. Nowhere is this context more important than in dementia assessments, where disease prevalence

varies significantly between different settings and reporting of key details is highly varied. By using diagnostic accuracy as the lens to investigate dementia diagnosis in primary care as has been done within this thesis, there is a risk of further contributing to this artificially narrow focus of test evaluation. The intention here was to use diagnostic accuracy as a device to highlight inconsistencies in dementia diagnosis in general practice, but also to 'reverse the perspective' and view diagnostic accuracy through the lens of dementia diagnosis in order to identify and discuss practical challenges for test evaluation in clinical settings. Both these aims have been successful, but have left gaps in understanding of wider issues around test evaluation such as the effects on patient health, clinician practice and joint decision making as discussed in depth by di Ruffano and colleagues[451] in their influential 2012 paper.

### 7.4.3. Limitations of the research process

An early plan of this research was to follow a cohort of patients over 12 months from the point of diagnosis in general practice to see who improved (indicating a false positive diagnosis) and who developed further symptoms or stayed the same (indicating a true positive diagnosis). This would have allowed assessment of a number of factors including discordant results, as well as investigation of other parts of the test process such as acceptability to the individual and to the clinician, GP referral and additional testing in secondary care.

Another plan was to conduct in-depth interviews with GPs in order to fully explore their choice of tests and testing decisions when using BCAs as part of the process for identifying dementia in clinical practice, but funding was not available to pay GPs for interview time. With reduced ambition and funding requirements, it was hoped to carry out a paid GP survey, as based on previous experiences[118] this may have secured a higher response rate than with an unpaid survey. Unfortunately there was insufficient funding available for this, and so the unpaid survey conducted was the clearest route for gathering a number of GPs' views on a range of questions relating to BCAs choice and use in practice. Focus groups with GPs could have been another option, and this may have exploited the group dynamic to elicit more responses and create discussion between respondents. This may also have resulted in less variation amongst responses and dominance of individual voices either between respondents themselves or between respondents and the moderator[458,459]. In addition, sample sizes are not guaranteed to have been higher or samples more representative than

they were in the survey. Response rates were relatively low for the survey (50% response rate - 52 practitioners responded, from a potential pool of at least 100 GPs) compared to previous surveys based on samples from a professional register which secured a 91% response rate[460] and a survey of GPs which explored how clinicians apply existing test accuracy metrics for diagnostic decision making which gained a 95% response rate[118]. Whilst low, the rate secured within this survey was in line with some other GP surveys distributed without cash or voucher incentives. One study characterizing response rates to mail-based surveys published in medical journals reported an average response rate of 54%[461]. A survey of Irish GP attitudes and practices in relation to screening, diagnosing, and disclosing a dementia diagnosis to patients also reported a response rate of useable surveys of 50%[19]. Another study looking at response rates of GPs and internal medicine specialists to combinations of mail survey followed up by a web-based survey, and vice versa, reported response rates of 63-71%[462]. Resource constraints and budget limitations meant that no financial incentives could be offered to potential contributors within a profession acknowledged to be short on time [134,136,373,463]. If resources had been available, financial incentives and the use of professional surveying providers such as Doctors.net.uk would have allowed a larger survey sample and may well have secured a higher number of responses such as those reported in the unpublished 2012 survey of 996 GPs across 5 countries conducted on behalf of Lilly[464]. Responses to this survey were not representative of GPs who were inactive on Twitter, were not members of the CHAIN network and were not known to the candidate's network of GPs who helped to distribute the survey. The distribution of responses showed fewer replies from the North East of England, Northern Ireland, the Scottish Highlands and Wales. Age of respondents and number of years in general practice was well distributed in line with national patterns of distribution[51]. Whilst it would have been valuable to ensure better general representation to the UK GP population and therefore greater generalisability, conclusions were tailored accordingly and designed not to make claims to broader generalisability of results than was justified within the data.

The rapid review of CPGs is likely to have missed some CPGs not published within the databases searched, as well as all relevant CPGs not published in English. This review was rapid by dint of time and resource allowed, and was a specific requirement

of examiners following the PhD viva in order to systematically investigate CPGs available for the use of BCAs as part of the process for identifying dementia in general practice. A future review would benefit from more detailed and thorough searches of more than 4 databases, as well as the greater procedural rigour introduced within the conduct of a full systematic review compared to a rapid review – namely the publication of the protocol as an open access peer reviewed article, searches including grey literature and forwards/backwards citation searching, and the carrying out of the review with a minimum of two reviewers conducting key processes (such as screening and data extraction) blindly and independently.

It would have been valuable to conduct a full review of how guideline advice matches cited supporting evidence, as in the process of this research it has been discovered that this cited evidence is not guaranteed to match guidance. This was not something recognised at the start of this project, and by the time it was realised quite how dissonant these elements can be it was too late to work this into the thesis without changing focus of the PhD significantly. This would be worth pursuing as a separate research project, and would continue and develop a theme that has emerged through this PhD thesis – that formal guidance is not always supported by best evidence.

The systematic review of direct comparisons of diagnostic accuracy focussed on two BCAs – MMSE and GPCOG. The rationale for this specific emphasis is clear, in that these were the two most commonly used tests identified within the overview of BCAs and the GP survey that was conducted, and the two most commonly quoted tests within guidelines. There are direct comparisons of diagnostic accuracy conducted with other tests, and an analysis including more BCAs would probably generate useful additional information. The main reasons behind the decision to limit the systematic review purely to two BCAs were that the justification for using these two tests due to their pre-eminence amongst existing evidence was strong, and there were limited resources to conduct a larger scale systematic review. It was judged to be better to carry out a specific and focused review looking at two BCAs, particularly using relatively new methods in a systematic review of direct comparisons of diagnostic accuracy, so the scope was restricted accordingly.

Finally, a clear limitation was that the primary focus was on reviews of diagnostic test accuracy, with other information (such as aspects of broader impact) reported as of secondary interest. It has become clear through the course of this PhD that whilst

accuracy is a useful measure for test evaluation, it is rarely the most important measure when considering the impact of a test and testing on patient related outcomes. In a future review, it would be valuable to explore analytical performance (e.g. validation, replication) of BCAs, as well as cost-effectiveness and broader impact (e.g. admin time, acceptability to patients and clinicians) within a specific, targeted review of the evidence. These factors are discussed in more detail in section 7.5.

### 7.4.4. Changes within the thesis

Little has been changed in the course of the PhD from the original plan stated in the registration documents and PhD midpoint transfer report. Within the overview, the only change between protocol and review was in the search strategy. It was intended to search an evidence database being developed by the University of Exeter Medical School. The database however was not complete at the time of searching and therefore searches were restricted to the priority databases.

Within the systematic review, there were two departures from the published protocol. One was regarding stratification of GPCOG where it had originally been stated that the most clinically relevant measures of GPCOG Patient, GPCOG Total and GPCOG Two stage (as GPCOG Informant has not been recommended for use by itself within a clinical setting[284]) would be used. Included studies reported four individual measures for GPCOG (Informant, Patient, Total and Two stage) so these four GPCOG subtypes were stratified, rather than across the pre-stated and arguably more clinically-relevant three subtypes.

The other departure from the protocol was in data extraction where it had been stated in the protocol that the data abstraction form would be piloted with two included studies, but in the event only one study was used to pilot the form, as it was a simple process and further trials were not needed.

Post-viva, the rapid review of CPGs was an additional piece of work conducted at the request of the examiners and has helped to asses and strengthen the evidence base and known context for conclusions around the availability of CPGs for this specific population and setting. Further work has also been carried out in triangulation of data from the survey, which has allowed a greater depth and context to be drawn from the findings within the wider context of published research literature and evidence identified elsewhere within this thesis. Recommendations

### 7.4.5. Background to the recommendations

New Research Framework proposals for Alzheimer's diagnosis from the National Institute for Aging and the Alzheimer's Association[465] (NIA-AA) provide an outline for defining Alzheimer's disease using biomarkers. This amounts to a fundamental reassessment of how we define Alzheimer's disease, moving from a clinical syndrome to a set of explicit biological features based upon neurofibrillary tangles and amyloid plaques. This new agenda has been positively welcomed by many institutes already active in the use of biomarkers. Such institutions have highlighted perceived benefits to clinical trials in driving improved biological profiling of trial participants, and the adoption of universal measures and language for the Alzheimer's research community. There are, however, fundamental questions still remaining around the exact nature of the disease, how biomarkers and clinical presentation interact and why many people with indicator biomarkers present may never develop clinical symptoms. A critique of this NIA-AA research framework for Alzheimer's disease by editors of the Cochrane Dementia and Cognitive Improvement Group (CDIG)[394] highlighted the danger of shifting diagnostic criteria and therefore disease boundaries from the research sphere into clinical practice, resulting in promoting the use of biomarkers within the clinic. This is already a feature of the latest NICE dementia guidelines, which support the use of CSF and imaging biomarkers to assist clinical diagnosis[383]. It has been indicated elsewhere in this thesis that not enough is known about the false positives and potential discordance between BCA results (i.e. if one person tests positive on test A but negative on test B, what are the underlying discrepancies?). What is emphasised in this critique of this new NIA-AA framework is that the potential for misclassification when using biomarkers as the sole diagnostic criteria is substantial, with ensuing implications of psychological, financial, legal and social costs for individuals and wider society[466]. Whilst these critiques of emergent biomarker technologies are compelling, they are not conclusive – and it may be that combinations of different test modalities using approaches such as staging, triaging or as add-ons to existing diagnostic pathways [467] may provide the much needed improvements to the early phases of the process for dementia identification within general practice[468].

A large volume of evidence was found around BCAs and their place in identifying dementia in general practice. This evidence is mixed in quality and findings, and there is a deficit of clear, robust evidence showing that BCAs have a high or acceptable

degree of accuracy and are fit for purpose. Considering all these weaknesses, we should ask: are BCAs the best methods for assessing cognitive performance in general practice? Are they actually cost effective, effective for patient reported outcome measures, effective for clinician reported outcome measures, with a clear benefit for clinical decision making and ongoing management decisions?

In addition to these questions, we should ask: what are BCAs? They are not brief, whatever brief means – many take well over ten minutes to complete in full, without allowing for set up, briefing and discussion of results. They are not always solely related to cognition, as many cover multiple domains which also incorporate other measures such as informant assessments and activities of daily living metrics. We also need better evidence on the role and value of informant reporting. Several BCAs such as the GPCOG, IQCODE and AD8[64,65,125,142,469] have sections for informants to complete, but the additional value in terms of other outcomes beyond cognitive function is currently underexplored.

As dementia is an umbrella term covering a range of disorders with varied presentations, only some of which are chiefly characterised by cognitive dysfunction, is cognition the one aspect of function to be focussing on within assessments? Or is a functional approach[470], as supported by the latest DSM-VI and ICD-11 guidelines, the future of dementia assessment? If we can address these questions rationally, we may find that there are other tools and assessments such as blood tests which are emerging as *at least* as effective in terms of diagnostic accuracy – and perhaps more acceptable to patients, clinicians, with improved cost-effectiveness. These alternatives may provide a great improvement on current available tests, and should not be discounted in the mistaken belief that current BCAs are adequate.

### 7.4.6. List of recommendations

From the research presented, 6 key recommendations are proposed:

**1.** More studies are needed which measure the diagnostic accuracy of BCAs and consistently address the general practice population, in terms of disease prevalence, sampling, recruitment, allocation, inclusion strategies and test administration. As detailed in Chapters 4 and 5, there is a paucity of studies carried out within a population relevant to general practice, using general practitioners to conduct assessment. Evidence identified within this thesis (detailed in Chapters 3,

4, 5 and 6) was broadly inconsistent in the way that participants were recruited, common groups found in general practice were excluded (such as people with cardiovascular problems or sensory impairments), general practitioners did not feature as assessors for the tests, and missing data were a significant problem.

**2.** The quality of overviews, systematic reviews and primary studies which assess the diagnostic accuracy of BCAs need to be improved in order to clearly and accurately report source data, employ clinically relevant BCAs, consistently use a study population that relates closely to the target population of interest, report basic accuracy data such as disease prevalence, and make the contents of the 2x2 table available to allow for further analysis as required. Evidence on the accuracy of BCAs disproportionately featured the MMSE, study populations did not relate consistently to general practice populations, BCA thresholds varied widely, and were not clearly reported in all cases. Sensitivity and specificity were not clearly reported and data were pooled in many cases. Variability found at the levels of study and systematic review was amplified within the overview.

**3.** To qualify as a BCA, a test should have the following 7 features:

i. Maximum 5 minutes administration time *including* time for instruction, interpretation and explanation – derived from guideline evidence presented in Chapter 3, and GP survey responses presented in Chapter 6;

ii. Free and simple to access – derived from guideline evidence presented in Chapter 3, clinically relevant outcomes presented in Chapter 5 and GP survey responses presented in Chapter 6;

iii. Validated in a general practice population – derived from evidence syntheses presented in Chapters 3, 4, and 5, clinically relevant outcomes presented in Chapter 5 and GP survey responses presented in Chapter 6;

iv. Can be administered within general practice – derived from evidence syntheses presented in Chapters 3, 4, and 5, and GP survey responses presented in Chapter 6;

v. Can be administered by practice staff who are not general practitioners – derived from GP survey responses presented in Chapter 6;

vi. Simple to use with clear instructions for interpretation – derived from clinically relevant outcomes presented in Chapter 5 and GP survey responses presented in Chapter 6;

vii. Free from education, language, and cultural bias – derived from guideline evidence presented in Chapter 3, evidence syntheses presented in Chapters 3, 4, and 5, clinically relevant outcomes presented in Chapter 5 and GP survey responses presented in Chapter 6.

**4.** There needs to be further assessment of other equally important factors for BCAs in primary care alongside diagnostic accuracy. These factors should certainly include acceptability to the patient; acceptability to the clinician; cost-effectiveness; test administration time; and possibly other factors. None of the evidence assessed within this thesis addresses the ability of tests to monitor performance over time. This was not a topic that was purposefully targeted within this PhD research, yet a number of GPs within the survey development and conduct referred to using BCAs to measure change or take a snapshot at different time points in order to measure decline or maintenance. Anecdotally, performance over time would be a useful measure for general practitioners and tests are already used in this way, yet there is little evidence on whether any of the current BCAs are useful for this purpose. There was little evidence found within primary and secondary sources on patient views, acceptability and understanding of BCAs. This is another significant gap which would benefit investigating as a priority. As patient and public involvement and shared decision making becomes more integrated within general practice, this lack of evidence will become more prominent and at odds with wider public health policy. Related to this, there was little evidence on GP views and GP understanding of testing in the context of BCA use as part of the process for identifying dementia within general practice. Whilst professional nihilism is often quoted as one of the barriers to improved dementia diagnosis rates within general practice[18,136,331-333,339,348,390], there may be many other reasons for missed diagnosis and misdiagnosis which have not yet been fully explored within this specific context.

More work should focus on this aspect of the testing process in the context of dementia diagnosis, particularly asking what test factors are considered most important and useful for GPs when first assessing potential dementia within the general practice setting. Specifically we need to understand how social and

interpersonal elements of the consultation may influence the GP's diagnostic decision making; what aspects of the diagnostic process are given greater weight (such as physiological symptoms; functional features such as Activities of Daily Living; or informant reports); and in the absence of some information (e.g. informant testimony), what other features are given greater prominence? Alongside understanding these factors better, there appears to be a corresponding gap in understanding the patient perspective during the diagnostic process. What factors are most valued by patients during diagnosis? What aspects are most problematic? If these more nuanced and complex elements of the diagnostic process can be better understood this will help to design processes that address some of the barriers in the system.

5. We need to interrogate the assumption that BCAs are the best measures to be using in general practice, and ask whether instead we should we be investing in different technologies and exploring the acceptability, feasibility and accuracy of other measures such as biomarkers. It may be that combinations of different test modalities using approaches such as staging, triaging or as add-ons to existing diagnostic pathways [467] may provide the much-needed improvements to the early phases of the process for dementia identification within general practice.

6. MMSE should not be the predominant BCA used either in research or in the clinic, and should not be considered as suitable for use as part of the process for identifying dementia in general practice. Whilst MMSE was prominent and the most popular BCA across many examples, the administration time of MMSE – where empirically established within a clinically relevant population - was found to be beyond the time allowed within standard general practice consultations. According to recent research[390], the use of MMSE as part of the assessment for potential dementia in general practice is *mandatory* in 12 European countries. Yet the ubiquity of MMSE may act as a barrier to progress; as long as it occupies a standard position of trust amongst BCAs within dementia diagnosis, there may be little progress in developing better tools.

Finally, it is important to return to the purpose of the test; where does it sit in the diagnostic pathway, and in the patient's healthcare journey? The BCA is used primarily as a triage tool within general practice, where an initial assessment is carried out and only patients with a positive test result are then referred on for further testing and work-

up in specialist memory clinics. If we are truly aiming to improve dementia diagnosis within general practice, it is imperative that accuracy be improved or new measures are found to replace current BCAs, which are not fit for purpose.

The Alzheimer's Society recently announced priorities for improving dementia research by 2025[471]. The second priority identified was to maximise the benefits of a dementia diagnosis to patients and their carers, and as part of this to research the acceptability, cost-effectiveness and health outcomes of innovations in diagnostics, including non-invasive tests that support making a diagnosis in primary care.

This thesis goes some way to addressing these aims by drilling down into the specific areas where current knowledge is lacking and shedding light to improve our understanding around the accuracy of BCAs within the context of general practice.

## 7.5. What does this PhD thesis add?

Through this PhD research, a number of important areas of uncertainty in the diagnostic accuracy of BCAs used as part of the process for identifying dementia within general practice have been identified. These uncertainties are often hidden within guidelines where it is assumed that much of the supporting evidence is grounded within general practice populations, with dementia prevalence similar to or the same as the real life populations where guidance and policies will be applied.

What has been discovered and demonstrated within this thesis is that much of the evidence assessing the diagnostic accuracy of BCAs when used as part of the process for identifying dementia in primary care is based upon studies with ill-suited populations, disease prevalence, sampling, recruitment, allocation, inclusion strategies and test administration.

Published CPGs relating specifically to BCA use in general practice are not common, and for those that were identified and analysed in the course of this PhD thesis, the recommendations are both inconsistent and not tailored for target populations. There is little information available on specific BCAs, the evidence behind recommendations and how CPGs are suited for culturally and linguistically diverse audiences. Key aspects of BCA selection and performance such as tailoring and disease prevalence are overlooked within current guidance, and there is a strong need to improve and broaden the evidence for CPG recommendations.

Very few studies assess BCAs using direct comparisons of diagnostic accuracy within the same study population. This is an issue as diagnostic accuracy assessments are open to a great deal of variation across a number of factors beyond variation in populations, settings, conditions of interest and administration. Test factors such as versions, assessors, testing environment, ordering and timings between testing all make a difference to how a test or tests perform. Reference standards are often imperfect and sometimes complex, made up of several tests (composite) or requiring a different administrative process compared to the index test. In many cases observed within the overview, reference standards were based upon one or more of the index tests being assessed. All this variation contributes in subtle and significant ways to the testing environment and may influence the results of a diagnostic accuracy assessment. Where two or more tests are directly compared within the same population, a number of these influences can be expected to affect both tests more-or-less equally, as they apply to the same participants in the same environment in the same way. One clear exception in this scenario is when there is missing data from one testing arm and not another, leading to an asymmetry in potential bias. This was one of the weaknesses identified within the systematic review looking at the accuracy of MMSE and GPCOG when directly compared to one another.

The paucity of high quality study level evidence designed and conducted specifically to assess the diagnostic accuracy of BCAs within a general practice population is a surprising and concerning finding. Whilst a lack of evidence was identified which allowed the robust assessment of diagnostic accuracy of BCAs with a general practice population, a great deal of evidence assessing BCA accuracy was found that was of average quality or poorly conducted and reported. This meant that not only were research findings highly limited in their usefulness, but the participants, finance and resources that had contributed to this evidence were ill served due to avoidable failures in conduct of the studies and reporting of the evidence.

Similarly, many of the evidence syntheses assessed were poorly suited for applying evidence to general practice due to methodological weaknesses, yet in many cases this incompatibility was difficult to identify. Even after having used established quality assessment tools such as QUADAS-2, the evidence needed unpicking in detail to reveal deficiencies in methodology and conduct such as missing data, inappropriate study designs and high degrees of incorporation bias.

Another discovery of this PhD research is the lack of evidence around the assessment of other test factors which impact on accuracy and test performance. These include acceptability to the patient, acceptability to the clinician, cost-effectiveness of the test and the testing process, test administration time, and test validity. Whilst these factors were not the primary focus of the systematic review and overview, instances within the evidence reviewed were positively identified where these measures were included alongside diagnostic accuracy evidence, and yet there were few examples where these data were reported. These factors beyond test accuracy have a strong influence on how testing improves health outcomes relative to the best alternative, such as no testing or another testing scenario [472]. Little systematically gathered evidence on these broader factors was found.

As detailed in Chapters 4 and 5, a number of studies which aimed to directly compare the diagnostic accuracy of two or more BCAs used as part of the process for identifying dementia in general practice fell short of this goal, through unclear or poor research methods, mixed methodological quality and poor reporting.

Many index tests were compared on an uneven basis, where the MMSE was also used at least in part as the reference standard within the CAMDEX or CAMCOG, leading to overestimation of effect and greater potential for incorporation bias[282]. The asymmetry of bias in this approach is a particular concern, as one index test (GPCOG) which does not feature within the reference standard would be disadvantaged when directly compared to another test (MMSE) which does. Many other biases observed within direct comparisons of diagnostic test accuracy assessments may affect both index tests relatively equally, as they influence the same population, setting and – broadly – administration process. Indeed, this is one of the stated benefits of direct comparisons of diagnostic accuracy over indirect comparison studies of diagnostic accuracy[473-475].

In summary, the thesis identifies and characterises these key issues:

- a lack of robust study level evidence specifically addressing a general practice population, in terms of disease prevalence, sampling, recruitment, allocation, inclusion strategies and test administration;
- The failure of evidence – where it does exist - to clearly and accurately convey source data, employ clinically relevant BCAs, consistently use a study population that relates closely to the target population of interest, report basic accuracy data

such as disease prevalence, and make the contents of the 2x2 table available to allow for further analysis as required;

- Ill-defined BCAs suitable for use in general practice; and
- A paucity of evidence on other equally important factors such as patient and clinician acceptability; cost-effectiveness; and test administration time for BCAs in primary care; alongside diagnostic accuracy.

Therefore the next step within this research would not be to conduct any more systematic reviews. It has been demonstrated that there are plenty of reviews into BCA accuracy, but the evidence is not particularly helpful in identifying which BCAs are most accurate or suitable for identifying dementia in general practice.

Instead, the suggestion would be to design and conduct a robust primary research comparative accuracy study to precisely address the weaknesses identified within this thesis. This study would also be designed to deliver detail on wider aspects of the testing process that impact on patient outcomes, such as clear empirical measurement of test administration time, acceptability to the patient, to clinicians and to wider society; cost-effectiveness; and ease of use. As part of this, a stakeholder reference group would be an essential feature of the research, with members made up of patients, relatives and clinicians who would be involved from inception in developing research questions, interrogating the data and disseminating the findings to greatest effect. The tools themselves would be clearly defined as BCA tools, being free at the point of use, with administration times suited to general practice application, possibly with an informant element, and independent from the reference standard. The GPCOG would be compared alongside one other test such as Mini-Cog, as although MMSE is the most prominent in research and (possibly) practice, it is not free and administration time has not reliably been shown to be suitable for general practice.

Alongside this study, work would be conducted to understand further what the general practitioner is actually trying to do within the diagnostic process for identifying dementia. Currently, the common perception is that there is a single, simple route to diagnosis[15,476] and this needs to be grounded in the reality of clinical consultation. It would be highly valuable to characterise in detail the variety of roles the GP is trying to fulfil here through the diagnostic process, from initial consultation through to assessment, referral and follow up. This would be conducted with general practitioners

on the research team, to ensure that the views of the clinicians were fully represented and built into the design, conduct and communication of the research.

In addition to the primary comparative accuracy study outlined above, it would be reasonable to conduct a trial of the effectiveness of BCAs when used to identify dementia in general practice with outcomes designed to translate clearly into benefits for patients[477]. This randomised controlled trial would involve a relevant stakeholder group from the start, with patients, relatives and clinicians involved meaningfully throughout the research from the design of research questions to the conduct of the trial and dissemination of research findings. Factors that the trial would measure include the direct effects of testing, using a clinical measure as the primary outcome, such as changes in clinical management (e.g. starting, stopping, or modifying treatment; ordering more tests; or watchful waiting) and decision-making, guided by these test results[472]. Assessments of the effects of testing on clinicians' diagnostic thinking and subsequent clinical decision making and management feature prominently in proposals for evaluation schemes of medical tests[308].

Secondary measures would consist of the indirect (non-clinical) effects of testing, in order to represent outcomes of relevance to patients, clinicians or other stakeholders such as emotional, behavioural, social or cognitive effects. It would be important to conduct scoping searches and focus groups with stakeholders to identify the most meaningful and relevant outcomes for people directly and indirectly affected by the process. In addition, an assessment of cost-effectiveness would be included in order to establish the actual costs of using individual BCAs as part of the process for identifying dementia in general practice. A model of the cost effectiveness of using BCAs within UK general practice[428] based upon a simulated cohort found that the use of MMSE, 6CIT or GPCOG would be a cost-effective strategy compared to clinical judgment within general practice. Another recent study looked at the cost-effectiveness of memory assessment services for the diagnosis and early support of patients with dementia in the UK[478]. The researchers found that diagnosis, treatment and follow-up care given to people with suspected dementia was effective but not cost-effective over the first six months after diagnosis. Another recent study of cost-effectiveness conducted in a German memory clinic[479] found a high range of costs and variation across diagnostic demands in this setting. There would be real value in

building upon these findings to assess cost-effectiveness alongside other patient and clinician relevant outcomes within a UK general practice setting.

The purpose of investigating the effectiveness and diagnostic value of BCAs for identifying dementia in general practice would be to objectively and systematically assess the suitability of BCAs as a tool for use in this population and setting. Whilst the suspicion may be that BCAs generally, or specific BCAs, may not be 'fit for purpose' in terms of identifying dementia within general practice, we still need to establish that this is the case using robust methodological and statistical approaches and reported in the fullest and most accurate ways possible. It is also important to establish what is meant by 'fit for purpose' in this context, as identifying dementia is not a straightforward process of GPs diagnosing a single discrete condition. It is likely that there is a far more complex problem being addressed here, and at present all the component parts of this process are grouped under the heading of dementia diagnosis.

What is needed is clear, specific, well designed primary research to begin to unpick these complexities and realistically address the challenges presented by the identification of dementia within general practice and primary care.

## 7.6. Reflections on the PhD

### 7.6.1. Evolution and development

The research ideas behind this PhD have been part of an ongoing conversation between me and Professor Chris Hyde, my Director of Studies, since at least 2013. The kernel of the research question came about as we considered the challenge of diagnosis in a condition such as dementia with uncertain aetiology, limited treatment options and initial presentation commonly at the level of general practice.

Following an unsuccessful application for an NIHR PhD Fellowship in the 2013/2014 funding round, the original research idea was revised from 20 interviews with GP practices and a cohort study of people initially diagnosed with dementia with a 12 month follow-up to a more modest research plan with scaled down ambition and making full use of existing research skills. This led to an application to the Alzheimer's Society Project Grant scheme for an overview, two systematic reviews, and a survey of GPs followed by in-depth interviews to explore issues arising from the previous

work. This application was rejected by the funders who predicted difficulties in involving GPs in research and ensuring clinical adoption of research findings.

Further refinement of the project plan proposed semi-structured interviews with a smaller sample of GPs, and this application was submitted to the Royal College of General Practitioners Scientific Foundation Board. Although positively received, the application was rejected on the basis of over ambition in the number of research objectives, and uncertainty around the impact of this research on clinical practice.

These comments directly influenced the final iteration of the research plan which scaled down ambition to two evidence syntheses and a survey of GPs to explore clinical reality of test selection and use, and to compare these views against the existing evidence base. Following post-viva requests from the Examiners, an additional piece of evidence synthesis work was conducted alongside further analysis of the survey results to triangulate findings with current evidence and evidence already identified within the PhD. This PhD research has been generously supported by the Diagnostics Theme of the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) South West Peninsula (PenCLAHRC), and the resources and associated support provided by this organisation has helped immeasurably in conducting this PhD thesis.

The judgements and comments received from all of these research boards and organisations reflect careful consideration by many experts within their areas who judge funding applications on a regular basis. This current research has been refined as a direct result of their careful consideration of proposals, and this work has evolved into a leaner, more efficient and well defined piece of research because of the time taken by each reviewer at the various stages described above. Whilst the process of securing support has not been straightforward and has - at times - proved challenging, this passage of refinement and evolution has proved ultimately beneficial to both the research project and my experience as an independent researcher.

## 7.7. Thanks and acknowledgements

### 7.7.1. Project Advisory Group

At the beginning of this venture, I had a strong sense that a Project Advisory Group would be a significant addition to my PhD studies. This view was based principally

upon previous research experience with Project and Research Advisory Groups which had been generally positive and beneficial to the work being undertaken.

My rationale was that not only would a range of expertise and experience help in shaping and interrogating my research decisions, but working with a group of 'critical friends' would force me to be clearer than I might otherwise be on motivations behind decisions, directions taken within the PhD, and to articulate these processes even at times when it felt uncomfortable to do so.

What I never anticipated was the level of engagement, generosity and tenacious goodwill I have experienced throughout the course of my studies. I am outstandingly grateful for the kindness I have been shown by Project Advisory Group members, shown in **Error! Reference source not found.**.

Table 22 Project Advisory Group Members

| |
|---|
| **Anne-Marie Boylan** - Departmental Lecturer & Senior Research Fellow NIHR CLAHRC Oxford, Nuffield Department of Primary Care Health Sciences, Medical Sciences Division at the University of Oxford. |
| **Nick Cartmell** - General Practitioner in Asburton, Devon. |
| **Teresa Dyer** - Expert by experience |
| **Willie Hamilton** - Professor of Primary Care Diagnostics at the University of Exeter Medical School and General Practitioner in Exeter, Devon. |
| **Ian McKeith** - Professor of Old Age Psychiatry, Newcastle, Theme lead of Newcastle NIHR Biomedical Research Unit in Lewy Body Dementias and Biomedical Research Centre in Ageing and Neurodegenerative conditions, Hon. Consultant in Old Age Psychiatry, Newcastle North Tyneside and Northumberland Mental Health NHS Trust; DLB International Consortium Lead |
| **Rupert McShane** - Associate Professor, Dementia Clinical Network Lead for the Oxford Academic Health Science Network, Consultant Old Age Psychiatrist, Coordinating Editor of the Cochrane Dementia & Cognitive Improvement Group. |
| **Yemisi Takwoingi** - Senior Research Fellow in Biostatistics at the University of Birmingham. |
| **Malcolm Turner** - Expert by experience. |

The eight individuals I approached were chosen as friends and colleagues of supervisors, members of other groups at the University of Exeter Medical School, people I have worked with previously and people whose work I have admired.

Everyone I asked to join my Project Advisory Group agreed with enthusiasm, and has stayed with me to the end. Not only have Project Advisory Group members remained steadfastly engaged and supportive of my undertakings, as individuals they have eagerly reviewed chapters of this thesis and provided many insightful and helpful comments along the way.

Having such a breadth of expert input from lay experts, old age psychiatrists, general practitioners, qualitative and quantitative research methodologists, research ethics specialists, diagnosticians, medical statisticians, epidemiologists, neuropsychologists and implementation specialists has enriched my understanding and examination of this area immeasurably. I hope the work I have produced within this thesis has justified their kind involvement, and that some of us may have the opportunity to work together again in years to come.

### 7.7.2. Supervisory team

I have been supported throughout my PhD research by an expert and engaged team of supervisors in Professor Chris Hyde, Professor of Public Health and Epidemiology at the University of Exeter Medical School (Director of Studies), Dr Mark Pearson, Senior Lecturer in Implementation Science & Knowledge Mobilisation at Hull York Medical School (secondary supervisor), and Dr David Llewellyn, Senior Research Fellow in Clinical Epidemiology at the University of Exeter Medical School (secondary supervisor). In particular, Professor Chris Hyde as my Director of Studies has given unstinting and good-humoured support throughout the project and beyond, providing links to international experts in the field and always making time to address issues and answer questions however menial. This opportunity to work with and learn from Professor Hyde has been a privilege as well as a sincere pleasure.

### 7.7.3. Wider networks

In addition to the enthusiastic and expert input I have received from my Project Advisory Group throughout my PhD research, I have been astonished by the amount of help I have received from the wider academic community.

Within the sometimes claustrophobic isolation of a PhD, the friendship, encouragement and willingness to help I have experienced from various academic sources has surpassed all expectations and provided many moments of respite from the morass of PhD scholarship. From Twitter support via the Shut Up and Write organisation (@SUWTUK) and various new networks using the #PhDForum and #PhDlife hashtags, to colleagues who have reviewed various drafts of papers, ethics applications, thesis chapters and conference presentations, I have been exceptionally fortunate in the support I have benefitted from throughout my research.

More widely, I have been moved by the number of researchers, authors and professionals from international organisations who have willingly provided assistance including individuals from the World Health Organisation, Alzheimer's Disease International (who shared a valuable pre-publication report), Eli Lilly (who shared unpublished research), the National Institutes of Aging in the United States of America and the Alzheimer's Society in the UK.

Individuals who have provided particular help include Becca Abbott, Yemisi Takwoingi, Louise Crathorne, Alison Bethel, Sanne van Kampen, Elzbieta Kuzma, Obi Ukoumunne, Jo Thompson Coon, Steph Powell, Jenny Lowe, Sue Whiffin, Becky Hardwick, Pim van den Dungen, Molly Wagster, and Jennifer Hill.

Finally, none of this would be possible without my husband John and our dog Kipper. Between them, they have kept me sane, exercised and fed and I'm inexpressibly grateful for our lives together.

## References


1. Alzheimer's Disease International. Dementia Statistics 2017. 2017; http://www.alz.co.uk/research/statistics.
2. Parkin E, Baker C. Dementia: policy, services and statistics. In: Library HoC. Vol 7007 17 October 2016. Westminster: House of Commons Library; 2016.
3. NICE. Dementia: Supporting people with dementia and their carers in health and social care. Clinical Guideline 422012.
4. Dubois B, Epelbaum S, Michon A, Funkiewiez A, Samri D, Hampel H. Screening For AD: Why And How? Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2014;10(4):P200.
5. Beck JC, Benson DF, Scheibel AB, Spar JE, Rubenstein LZ. Dementia in the elderly: the silent epidemic. Annals of Internal Medicine. 1982;97(2):231-241.
6. Larson EB, Langa KM. What's the "Take Home" from Research on Dementia Trends? PLoS medicine. 2017;14(3):e1002236.
7. Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G. Timely diagnosis for Alzheimer's disease: A literature review on benefits and challenges. Journal of Alzheimer's Disease. 2016;49(3):617-631.
8. Brayne C. Interpretation of dementia diagnosis and treatment trends in the UK over time. The Lancet Public Health. 2017;2(3):e128-e129.
9. Burns A. A new dementia currency in primary care. In: England N. https://www.england.nhs.uk/blog/alistair-burns-18/2016.
10. Chithiramohan A, Iliffe S, Khattak I. Identifying barriers to diagnosing dementia following incentivisation and policy pressures: General practitioners' perspectives. Dementia. 2016:1471301216682625.
11. Bell S, Harkness K, Dickson J, Blackburn D. A diagnosis for£ 55: what is the cost of government initiatives in dementia case finding. Age and ageing. 2015;44(2):344-345.
12. Alzheimer Europe. Alzheimer Europe: Five-Country Survey of Carers Highlights Continuing Delays in Dementia Diagnosis Across Countries2017.
13. Chambers LW, Sivananthan S, Brayne C. Is Dementia Screening of Apparently Healthy Individuals Justified? Advances in preventive medicine. 2017;2017.
14. APPGD. Dementia Rarely Travels Alone: Living with Dementia and Other Conditions. London: House of Commons; 2016.
15. Walker IF, Lord PA, Farragher TM. Variations in dementia diagnosis in England and association with general practice characteristics. Primary Health Care Research & Development. 2017:1-7.
16. Dodd E, Cheston R, Cullum S, et al. Primary care-led dementia diagnosis services in South Gloucestershire: Themes from people and families living with dementia and health care professionals. Dementia. 2015;15(6):1586-1604.
17. Caruana-Pulpan O, Scerri C. Practices in diagnosis, disclosure and pharmacotherapeutic management of dementia by general practitioners–a national survey. Aging & mental health. 2014;18(2):179-186.
18. Cahill S, Clark M, O'Connell H, Lawlor B, Coen RF, Walsh C. The attitudes and practices of general practitioners regarding dementia diagnosis in Ireland. Int J Geriatr Psychiatr. 2008;23.
19. Cahill S, Clark, M., Walsh, C., O'Connell, H., & Lawlor, B. Dementia in primary care: the first survey of Irish general practitioners. International Journal of Geriatric Psychiatry. 2006;21(4):319-324.
20. Brodaty H, Howarth GC, Mant A, Kurrle SE. General practice and dementia. A national survey of Australian GPs. Med J Aust. 1994;160.
21. Brækhus A, Engedal K. Diagnostic work-up of dementia: a survey among Norwegian general practitioners. Brain Aging. 2002;2(4):63-67.
22. Prince M, Bryce R, Ferri C. World Alzheimer Report 2011: The benefits of early diagnosis and intervention: Alzheimer's Disease International; 2011.

**23.** James Lind Alliance. Outcomes of the James Lind Alliance Dementia priority setting partnership 2013.
**24.** Waldemar G, Phung KT, Burns A, et al. Access to diagnostic evaluation and treatment for dementia in Europe. International journal of geriatric psychiatry. . 2007;22(1):47-54.
**25.** Aminzadeh F, Molnar F, Dalziel WB, Mounde C, Ayotte D. A Scoping Interpretive Review of Literature on Perspectives and Practices of Primary Care Physicians Vis-à-vis Diagnosis and Management of Community Living Older Persons with Dementia. . Regional Geriatric Program of Eastern Ontario. 2012a.
**26.** ADI. National Dementia Plans: Early Detection and Diagnosis. Unpublished: Alzheimer's Disease International; 2017.
**27.** Robinson L, Tang E, Taylor J. Dementia: timely diagnosis and early intervention. BMJ. 2015;350.
**28.** Alzheimer's Society. Dementia 2012: A National Challenge. . London 2012.
**29.** Aminzadeh F, Molnar FJ, Dalziel WB, Ayotte D. A Review of Barriers and Enablers to Diagnosis and Management of Persons with Dementia in Primary Care. Canadian Geriatrics Journal. 2012;15(3):85.
**30.** Koch T, Iliffe S. Rapid appraisal of barriers to the diagnosis and management of patients with dementia in primary care: a systematic review. . BMC Family Practice. 2010;11(1):52.
**31.** Le Couteur D, Doust J, Creasey H, Brayne C. Political drive to screen for predementia: not evidence based and ignores the harms of diagnosis. . BMJ. 2013;347(7925):f5125.
**32.** Jessen F. Re: Political drive to screen for pre-dementia: not evidence based and ignores the

harms of diagnosis. 20 September 2013 in response to Le Couteur, D. G. L., Doust, J., Creasey, H., & Brayne, C. Political drive to screen for pre-dementia: not evidence based and ignores the harms of diagnosis. . BMJ. 2013;34.
**33.** ALCOVE. Timely Diagnosis of Dementia [WP5] 2013.
**34.** Dementia. A-PPGO. Unlocking diagnosis: House of Commons; 2012.
**35.** Watson R, Bryant J, Sanson-Fisher R, Mansfield E, Evans T-J. What is a 'timely'diagnosis? Exploring the preferences of Australian health service consumers regarding when a diagnosis of dementia should be disclosed. BMC health services research. 2018;18(1):612.
**36.** McLeanm S. Assessing dementia. Part II: Clinical, functional, neuropsychological and social issues. Australian and New Zealand journal of psychiatry. 1987;21(3):284-304.
**37.** Sachdev PS, Blacker D, Blazer DG, et al. Classifying neurocognitive disorders: the DSM-5 approach. Nat Rev Neurol. 2014;10(11):634-642.
**38.** Sathyanarayana Rao TS, Jacob KS, Shaji KS, et al. Dementia and the International Classification of Diseases-11 (Beta Version). Indian Journal of Psychiatry. 2017;59(1):1-2.
**39.** De Lepeleire J, Wind A, Iliffe S, et al. The primary care diagnosis of dementia in Europe: an analysis using multidisciplinary, multinational expert groups. Aging and Mental Health. 2008;12(5):568-576.
**40.** Pathways N. Dementia diagnosis and assessment. Developed from Dementia (2006) NICE guideline CG42. [Accessed online 05/05/2015: http://pathways.nice.org.uk/pathways/dementia/dementia-diagnosis-and-assessment#content=view-node%3Anodes-diagnosis-and-assessment] 2012. http://pathways.nice.org.uk/pathways/dementia/dementia-diagnosis-and-assessment#content=view-node%3Anodes-diagnosis-and-assessment. Accessed 05/05/2015.
**41.** Woodford HJ, George J. Cognitive assessment in the elderly: a review of clinical methods. QJM: An International Journal of Medicine. 2007;100(8):469-484.

**42.** Velayudhan L, Ryu S-H, Raczek M, et al. Review of brief cognitive tests for patients with suspected dementia. International psychogeriatrics. 2014;26(08):1247-1262.

**43.** Culverwell A, Milne A, Guss R, Tuppen J. Screening for dementia in primary care: how is it measuring up? Quality in Ageing and Older Adults. 2008;9(3):39-44.

**44.** Kansagara D, Freeman M. A systematic evidence review of the signs and symptoms of dementia and brief cognitive tests available in VA: Department Veterans Affairs, Veterans Health Administration, Health Services Research & Development Service; 2010.

**45.** Boustani M, Peterson B, Hanson L, Harris R, Lohr KN. Screening for dementia in primary care: a summary of the evidence for the U.S. preventive services task force. Ann Intern Med. 2003;138.

**46.** Burns A, Buckman L. Timely diagnosis of dementia: integrating perspectives, achieving consensus. London: BMA & NHS England. 2013.

**47.** Knapp M, Comas-Herrera A, Somani A, Banerjee S. Dementia: international comparisons. LSE PSSRU: National Audit Office; 2007.

**48.** US Congressional Committee on Aging. Alzheimer's Disease and dementia: A Comparison of International Approaches. Senate: US Government Information; 2012.

**49.** Prins A, Hemke F, Pols J, van Charante EPM. Diagnosing dementia in Dutch general practice: a qualitative study of GPs' practices and views. Br J Gen Pract. 2016;66(647):e416-e422.

**50.** Iliffe S, Manthorpe J, Eden A. Sooner or later? Issues in the early diagnosis of dementia in general practice: a qualitative study. . Family Practice. 2003;20(4):376-381.

**51.** England PH. National General Practice Profiles. 2016; This spine chart provides a summary of practice demography, deprivation, patient satisfaction and life expectancy estimates. Available at: http://fingertips.phe.org.uk/profile/general-practice/data#mod,1,pyr,2015,pat,19,par,-,are,-,sid1,2000005,ind1,-,sid2,-,ind2,-. Accessed 01/03/2016, 2016.

**52.** Downs MG. The role of general practice and the primary care team in dementia diagnosis and management. Int J Geriatr Psychiatry. 1996;11.

**53.** Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ. 2012;344.

**54.** Carrillo MC, Brashear HR, Logovinsky V, et al. Can we prevent Alzheimer's disease? Secondary "prevention" trials in Alzheimer's disease. Alzheimer's & Dementia. 2013;9(2):123-131. e121.

**55.** Prince M, Knapp M, Guerchet M, et al. Dementia UK: Update (2nd Edition) 2014.

**56.** Alzheimer's Association. Alzheimer's disease facts and figures. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2016;12(4):459-509.

**57.** OECD. Addressing Dementia: OECD Publishing; 2015.

**58.** POST. New Drugs for Dementia; No. 535 2016.

**59.** Rollo JL, Banihashemi N, Vafaee F, Crawford JW, Kuncic Z, Holsinger RD. Unraveling the mechanistic complexity of Alzheimer's disease through systems biology. Alzheimer's & Dementia. 2016;12(6):708-718.

**60.** APPGD. Unlocking dementia: The key to improving the lives of people with dementia 2012.

**61.** Barrett E, Burns A. Dementia Revealed: What Primary Care Needs to Know. A Primer for General Practice 2014.

**62.** NIH. Assessing Cognitive Impairment in Older Patients: A Quick Guide for Primary Care Physicians. In: Center AsDEaR. https://www.nia.nih.gov/alzheimers/publication/assessing-cognitive-impairment-older-patients2014.

**63.** DHHS. Providing the Annual Wellness Visit (AWV). Medicare Preventative Services.: Department of Health and Human Services: Centers for Medicare and Medicaid Services.; 2012.

**64.** Cordell CB, Borson S, Boustani M, et al. Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. Alzheimer's & Dementia. 2013;9(2):141-150.

**65.** Ballard C, Burns A, Corbett A, Livingston G, Rasmussen J. Helping you to assess cognition: A practical toolkit for clinicians. The Alzheimer's Society. [Accessed online 05/05/2015 http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2532 ]. 2015. http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2532. Accessed 05/05/2015.

**66.** GSA. KAER Toolkit: 4-Step Process to Detecting Cognitive Impairment and Earlier Diagnosis of Dementia 2017.

**67.** Graham J. New Toolkits Help Physicians Detect, Diagnose, and Manage Dementia. Jama. 2017.

**68.** Boustani M, Peterson B, Hanson L, Harris R, Lohr KN. Screening for dementia in primary care: a summary of the evidence for the US Preventive Services Task Force. Annals of internal medicine. 2003;138(11):927-937.

**69.** Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and Delayed Diagnosis of Dementia in Primary Care: Prevalence and Contributing Factors. Alzheimer disease and associated disorders. 2009;23(4):306-314.

**70.** Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging & mental health. 2011;15(6):978-984.

**71.** Valcour VG, Masaki KH, Curb JD, Blanchette PL. The detection of dementia in the primary care setting. Archives of Internal Medicine. 2000;160(19):2964-2968.

**72.** Wilkinson D, Stave C, Keohane D, Vincenzino O. The role of general practitioners in the diagnosis and treatment of Alzheimer's disease: a multinational survey. J Int Med Res. 2004;32.

**73.** Brooker D LFJ, Evans S, Karim S,. Timely Diagnosis of Dementia: Synthesis Report 2013.

**74.** Bossuyt P, Leeflang M. Chapter 6: Developing Criteria for Including Studies. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Vol 42008.

**75.** Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Cochrane handbook for systematic reviews of diagnostic test accuracy. Version 0.9. 0. London: The Cochrane Collaboration. 2010.

**76.** Hofmann B, Welch HG. New diagnostic tests: more harm than good. BMJ: British Medical Journal (Online). 2017;358.

**77.** McInnes MD, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. Jama. 2018;319(4):388-396.

**78.** Nuffield Council on Bioethics. Dementia: ethical issues 2009.

**79.** Ashford JW, Borson S, O'Hara R, et al. Should older adults be screened for dementia? It is important to screen for evidence of dementia! Alzheimer's & Dementia. 2007;3(2):75-80.

**80.** World Health Organisation. Dementia. Factsheet No.362. 2016. http://www.who.int/mediacentre/factsheets/fs362/en/. Accessed 28/04/2017.

**81.** Mate KE, Magin PJ, Brodaty H, et al. An evaluation of the additional benefit of population screening for dementia beyond a passive case-finding approach. International journal of geriatric psychiatry. 2017;32(3):316-323.

**82.** National Screening Committee. UK NSC dementia screening recommendation. online 2015.

**83.** Committee UNS. UK NCS dementia screening recommendation. January 20152015

**84.** Borson S, Frank L, Bayley PJ, et al. Improving dementia care: the role of screening and detection of cognitive impairment. Alzheimer's & Dementia. 2013;9(2):151-159.

**85.** ARUK. Population and targeted case finding: policy statement2016.

**86.** Bonita R, Beaglehole R, Kjellström T. Basic epidemiology: World Health Organization; 2006.
**87.** Moyer VA, on behalf of the USPSTF. Screening for cognitive impairment in older adults: U.s. preventive services task force recommendation statement. Annals of Internal Medicine. 2014;160(11):791-797.
**88.** Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: A systematic review for the u.s. preventive services task force. Annals of Internal Medicine. 2013;159(9):601-612.
**89.** Research USDoVAHS, Service D, Kansagara D, Freeman M. A systematic evidence review of the signs and symptoms of dementia and brief cognitive tests available in VA: Department Veterans Affairs, Veterans Health Administration, Health Services Research & Development Service; 2010.
**90.** UK National Screening Committee. UK NCS dementia screening recommendation. January 20152015
**91.** Robinson L, Tang E, Taylor J-P. Dementia: timely diagnosis and early intervention. BMJ : British Medical Journal. 2015;350.
**92.** Livingston G, Baio G, Sommerlad A, et al. Effectiveness of an intervention to facilitate prompt referral to memory clinics in the United Kingdom: Cluster randomised controlled trial. PLOS Medicine. 2017;14(3):e1002252.
**93.** Philips E, Walters A, Biju M, Kuruvilla T. Population-based screening for dementia: controversy and current status. Progress in Neurology and Psychiatry. 2016;20(1):6-10.
**94.** Department of Health. Living Well With Dementia: A National Dementia Strategy2011.
**95.** Rasmussen J. Back to Back: General practitioners should be conducting targeted screening for dementia in people aged 65 to 74: Yes. Journal of Primary Health Care. 2014;6(3):245-247.
**96.** Oxford University Press. Aetiology (definition). Oxford English Dictionary. Online: Oxford University Press; 2017.
**97.** Starr JM, Walesby KE. Diagnosis and management of dementia in older people. Medicine. 2017;45(1):51-54.
**98.** Iliffe S. Back to Back: General practitioners should be conducting targeted screening for dementia in people aged 65 to 74: No. Journal of Primary Health Care. 2014;6(3):247-249.
**99.** Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: a systematic review for the US Preventive Services Task Force. Annals of internal medicine. 2013;159(9):601-612.
**100.** Seitz DP, Fage BA, Chan CCH, et al. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a primary care setting. Cochrane Database of Systematic Reviews. 2014(12).
**101.** Lees RA, Stott DJ, McShane R, Noel-Storr AH, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings. Cochrane Database of Systematic Reviews. 2014b(10).
**102.** Hendry K, Lees Rosalind A, McShane R, Noel-Storr Anna H, Stott David J, Quinn Terry J. AD-8 for diagnosis of dementia across a variety of healthcare settings. Cochrane Database of Systematic Reviews. 2014(5). http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011121/abstracthttp://onlinelibrary.wiley.com/store/10.1002/14651858.CD011121/asset/CD011121.pdf?v=1&t=idlm4c21&s=8dafa435a321ce12c678db51152a73d7cc4fcb37.
**103.** Harrison JK, Fearon P, Noel-Storr A, McShane R, Stott D, Quinn T. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting.

. Cochrane Database of Systematic Reviews 2014(7).

**104.** Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. Cochrane Database of Systematic Reviews. 2015(10).
**105.** Creavin ST, Wisniewski S, Noel-Storr AH, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. Cochrane Database of Systematic Reviews. 2016(1).
**106.** Arevalo-Rodriguez I, Smailagic N, Roque IFM, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). The Cochrane database of systematic reviews. 2015(3):Cd010783.
**107.** Society As. Dementia 2014: Opportunity for change.

. 2014.
**108.** van Hout H, Vernooij-Dassen M, Bakker K, Blom M, Grol R. General practitioners on dementia: tasks, practices and obstacles. Patient Education and Counseling.39(2):219-225.
**109.** NICE Pathways. Dementia diagnosis and assessment. 2016:10.
**110.** Gauthier S, Patterson C, Chertkow H, et al. Recommendations of the 4th Canadian Consensus Conference on the Diagnosis and Treatment of Dementia (CCCDTD4). Canadian geriatrics journal. 2012;15(4):120-126.
**111.** Guideline Adaptation Committee. Clinical Practice Guidelines and Principles of Care for People with Dementia.

. In: Committee GA. Sydney. 2016.
**112.** SIGN. Management of patients with dementia: A National Clinical Guideline. Edinburgh, Scotland: SIGN; 2006.
**113.** Bradford A, Kunik, M. E., Schulz, P., Williams, S. P., & Singh, H. . Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. . Alzheimer disease and associated disorders. 2009;23(4):306.
**114.** Brayne C. Interpretation of dementia diagnosis and treatment trends in the UK over time. The Lancet Public Health.2(3):e128-e129.
**115.** Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research. 1975;12(3):189-198.
**116.** Ismail Z, Rajji TK, Shulman KI. Brief cognitive screening instruments: an update. International Journal of Geriatric Psychiatry. 2010;25(2):111-120.
**117.** Lorentz WJ, Scanlan JM, Borson S. Brief screening tests for dementia. The Canadian Journal of Psychiatry. 2002;47(8):723-733.
**118.** Davenport C. Systematic reviews and meta-analyses of test accuracy: developing methods that meet practitioners' needs. University of Birmingham Research Archive e-theses repository: University of Birmingham; 2012.
**119.** Brodaty H, Connors MH, Loy C, et al. Screening for Dementia in Primary Care: A Comparison of the GPCOG and the MMSE. Dementia and Geriatric Cognitive Disorders. 2016;42(5-6):323-330.
**120.** Tsoi KK, Chan JY, Hirai HW, Wong SY, Kwok TC. Cognitive tests to detect dementia: a systematic review and meta-analysis. JAMA internal medicine. 2015;175(9):1450-1458.
**121.** Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Systematic reviews. 2012;1(1):10.
**122.** Tricco AC, Langlois E, Straus SE, Organization WH. Rapid reviews to strengthen health policy and systems: a practical guide: World Health Organization; 2017.
**123.** Polisena J, Garritty C, Kamel C, Stevens A, Abou-Setta AM. Rapid review programs to support health care and policy decision making: a descriptive analysis of processes and methods. Systematic reviews. 2015;4(1):26.

**124.** Dobbins M. Rapid review guidebook steps for conducting a rapid review. Hamilton:| Resource Details| National Collaborating Centre for Methods and Tools. 2017.

**125.** Woodford H, George J. Cognitive assessment in the elderly: a review of clinical methods. Qjm. 2007;100(8):469-484.

**126.** Schiffer RB, Slater RJ. Neuropsychiatric Features of Multiple Sclerosis Recognition and Management. presented at: Seminars in Neurology1985.

**127.** Billig N, Ahmed SW, Kenmore P, Amaral D, Shakhashiri MZ. Assessment of depression and cognitive impairment after hip fracture. Journal of the American Geriatrics Society. 1986;34(7):499-503.

**128.** Kiernan RJ, Mueller J, LANGSTON JW, Van Dyke C. The Neurobehavioral Cognitive Status Examination: A brief but differentiated approach to cognitive assessment. Annals of internal medicine. 1987;107(4):481-485.

**129.** Jacova C, Kertesz A, Blair M, Fisk JD, Feldman HH. Neuropsychological testing and assessment for dementia. Alzheimer's & Dementia. 2007;3(4):299-317.

**130.** Milne A, Culverwell A, Guss R, Tuppen J, Whelton R. Screening for dementia in primary care: a review of the use, efficacy and quality of measures. International Psychogeriatrics. 2008;20(05):911-926.

**131.** Menon R, Larner AJ. Use of cognitive screening instruments in primary care: the impact of national dementia directives (NICE/SCIE, National Dementia Strategy). Family practice. 2010:cmq100.

**132.** Deveugele M, Derese A, van den Brink-Muinen A, Bensing J, De Maeseneer J. Consultation length in general practice: cross sectional study in six European countries. Bmj. 2002;325(7362):472.

**133.** Gray DP, Orton P. Consultation length. British Journal of General Practice. 2017;67(656):108-109.

**134.** Hobbs FR, Bankhead C, Mukhtar T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. The Lancet. 2016;387(10035):2323-2330.

**135.** RCGP. It's Your Practice. A patient guide to GP services. 2011. https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&ved=0ahUKEwj1q86svb_TAhXILsAKHVDlCkEQFghHMAY&url=https%3A%2F%2Fwww.nhs.uk%2FchoiceintheNHS%2FYourchoices%2FGPchoice%2FDocuments%2Frcgp_iyp_full_booklet_web_version.pdf&usg=AFQjCNF5L9r3hp1BEgTd9fsZk4cBdC43BA&sig2=lXPDXCN2UHa8y-xdIi_SXg&cad=rja.

**136.** Croxson CH, Ashdown HF, Hobbs FR. GPs' perceptions of workload in England: a qualitative interview study. British Journal of General Practice. 2017;67(655):e138-e147.

**137.** Larner AJ. Speed versus accuracy in cognitive assessment when using CSIs. Progress in Neurology and Psychiatry. 2015;19(1):21-24.

**138.** Benbow SM, Jolley D, Greaves IC. Improving diagnosis of dementia in primary care. Progress in Neurology and Psychiatry. 2015;19(1):4-4.

**139.** Lees R, Selvarajah J, Fenton C, et al. Test Accuracy of Cognitive Screening Tests for Diagnosis of Dementia and Multidomain Cognitive Impairment in Stroke. Stroke. 2014;45(10):3008-3018.

**140.** Scheltens P, Rockwood K. How golden is the gold standard of neuropathology in dementia? Alzheimer's & Dementia: The Journal of the Alzheimer's Association.7(4):486-489.

**141.** Mitchell AJ, Malladi S. Screening and case finding tools for the detection of dementia. Part I: evidence-based meta-analysis of multidomain tests. The American Journal of Geriatric Psychiatry. 2010;18(9):759-782.

**142.** Brodaty H, Pond D, Kemp NM, et al. The GPCOG: a new screening test for dementia designed for general practice. Journal of the American Geriatrics Society. 2002;50(3):530-534.

**143.** Kirby M, Denihan A, Bruce I, Coakley D, Lawlor BA. The clock drawing test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly. International Journal of Geriatric Psychiatry. 2001;16(10):935-940.

**144.** Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. Int J Geriatr Psychiatry. 2000;15(11):1021-1027.

**145.** Tokuhara KG, Valcour VG, Masaki KH, Blanchette PL. Utility of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for dementia in a Japanese-American population. Hawaii medical journal. 2006;65(3):72-75.

**146.** Hooijer C, Dinkgreve M, Jonker C, Lindeboom J, Kay D. Short screening tests for dementia in the elderly population. I. A comparison between AMTS, MMSE, MSQ and SPMSQ. International journal of geriatric psychiatry. 1992;7(8):559-571.

**147.** Brooke P, Bullock R. Validation of a 6 item cognitive impairment test with a view to primary care usage. International journal of geriatric psychiatry. 1999;14(11):936-940.

**148.** Naqvi RM, Haider S, Tomlinson G, Alibhai S. Cognitive assessments in multicultural populations using the Rowland Universal Dementia Assessment Scale: a systematic review and meta-analysis. Canadian Medical Association Journal. 2015;187(5):E169-E175.

**149.** Buschke H, Kuslansky G, Katz M, et al. Screening for dementia with the memory impairment screen. Neurology. 1999;52(2):231-231.

**150.** Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. Neuropsychiatry Neuropsychol Behav Neurol. 1988;1(2):111-117.

**151.** Tangalos EG, Smith GE, Ivnik RJ, et al. The Mini-Mental State Examination in general medical practice: clinical utility and acceptance. presented at: Mayo Clinic Proceedings1996.

**152.** Meulen E, Schmand B, Van Campen J, et al. The seven minute screen: a neurocognitive screening test highly sensitive to various types of dementia. Journal of Neurology, Neurosurgery & Psychiatry. 2004;75(5):700-705.

**153.** de Yébenes MJG, Otero A, Zunzunegui MV, Rodríguez-Laso A, Sánchez-Sánchez F, Del Ser T. Validation of a short cognitive tool for the screening of dementia in elderly people with low educational level. International journal of geriatric psychiatry. 2003;18(10):925-936.

**154.** Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. International journal of geriatric psychiatry. 2006;21(11):1078-1085.

**155.** Lischka AR, Mendelsohn M, Overend T, Forbes D. A systematic review of screening tools for predicting the development of dementia. Canadian Journal on Aging/La Revue canadienne du vieillissement. 2012;31(03):295-311.

**156.** Yokomizo JE, Simon SS, de Campos Bottino CM. Cognitive screening for dementia in primary care: a systematic review. . International Psychogeriatrics. 2014;26((11)):1783-1804.

**157.** Lees RA, Stott DJ, McShane R, Noel-Storr AH, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings. Cochrane Database of Systematic Reviews. 2014a(10).

**158.** Arevalo-Rodriguez I, Segura O, Solà I, Bonfill X, Sanchez E, Alonso-Coello P. Diagnostic tools for alzheimer's disease dementia and other dementias: an overview of diagnostic test accuracy (DTA) systematic reviews. BMC neurology. 2014;14(1):183.

**159.** Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society open science. 2014;1(3):140216.

**160.** Oxford University Press. Guideline (definition). 2017; https://en.oxforddictionaries.com/definition/guideline.

**161.** Lohr KN, Field MJ. Guidelines for clinical practice: from development to use: National Academies Press; 1992.

**162.** NICE. Types of guideline. 2017; https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/types-of-guideline.

**163.** Department of Health Ireland. The Irish National Dementia Strategy 2014.

**164.** Di Fiandra T, Canevelli M, Di Pucchio A, Vanacore N. The Italian Dementia National Plan. Annali dell'Istituto Superiore di Sanità. 2015;51:261-264.

**165.** Alzheimer Portugal. Intervenção alzheimer: Trabalho preparatório para a conferência "doença de Alzheimer: Que políticas" 2009.

**166.** Department of Health. Improving dementia services in Northern Ireland - a regional strategy. In: Department of Health NI. https://www.health-ni.gov.uk/publications/improving-dementia-services-northern-ireland-regional-strategy2011.

**167.** Scottish Government. Scotland's National Dementia Strategy (2013 - 2016). http://www.scotland.gov.uk/Resource/0042/00423472.pdf2012.

**168.** Welsh National Assembly. National Dementia Vision for Wales: Dementia Supportive Communities. http://www.alzheimer-europe.org/content/download/20741/152710/file/National%20Dementia%20Vision%20for%20Wales.pdf2011.

**169.** Alzheimer Europe. National Dementia Strategies. 2017; http://www.alzheimer-europe.org/Policy-in-Practice2/National-Dementia-Strategies.

**170.** Brodsky J. Addressing Alzheimer's And Other Types of Dementia: Israeli National Strategy. In: Health Mo. http://www.alzheimer-europe.org/content/download/117195/734185/file/Israel%20national%20dementia%20strategy%202013.pdf2013.

**171.** Galvin JE, Sadowsky CH. Practical guidelines for the recognition and diagnosis of dementia. The Journal of the American Board of Family Medicine. 2012;25(3):367-382.

**172.** Field MJ, Lohr KN. Clinical practice guidelines: directions for a new program: National Academies Press; 1990.

**173.** Health NIf, Excellence C. Dementia: Assessment, management and support for people living with dementia and their carers: National Institute for Health and Care Excellence London; 2018.

**174.** Committee GA. Clinical practice guidelines and principles of care for people with dementia. Sydney: Guideline Adaptation Committee. 2016.

**175.** Kennedy HDP. 3 rd Canadian Consensus Conference on Diagnosis and Treatment of Dementia. 2006.

**176.** Excellence SCIf. Practice Guide 15. Dignity in care: SCIE London, UK; 2006.

**177.** Luning-Koster M, Perry M, van Charante Moll E, Vernooij-Dassen M, Wiersma T, Burgers JS. Summary of Dutch College of General Practitioners'(NHG) practice guideline'Dementia'. Nederlands tijdschrift voor geneeskunde. 2012;156(49):A5323-A5323.

**178.** Boomsma L, Boukes F, Wind A, Assendelft W. Summary of the practice guideline'Dementia'(second revision) from the Dutch College of General Practitioners. Nederlands tijdschrift voor geneeskunde. 2004;148(24):1191-1197.

**179.** Huppert FA, Cabelli ST, Matthews FE. Brief cognitive assessment in a UK population sample–distributional properties and the relationship between the MMSE and an extended mental state examination. BMC geriatrics. 2005;5(1):7.

**180.** Hobson J. The Montreal Cognitive Assessment (MoCA). Occupational Medicine. 2015;65(9):764-765.

**181.** Iliffe S, Robinson, L., Brayne C, Goodman, C., Rait G, Manthorpe J, Ashley P. Primary care and dementia: 1. diagnosis, screening and disclosure. Int. J. Geriat. Psychiatry. 2009;24:895–901.

**182.** D G, M D, M V-D, N S. Stigma and GPs' perceptions of dementia. Aging and Mental Health. 2015;20(4):391-400.

**183.** Bradford A, Kunik, M. E., Schulz, P., Williams, S. P.,, Singh H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. Alzheimer disease and associated disorders. 2009;23(4):306.
**184.** Koch T, Iliffe S. Rapid appraisal of barriers to the diagnosis and management of patients with dementia in primary care: a systematic review. BMC Family Practice. 2010;11(1):52.
**185.** Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging & mental health. 2011;15(6):978-984.
**186.** Aminzadeh F, Molnar FJ, Dalziel WB, et al. A Review of Barriers and Enablers to Diagnosis and Management of Persons with Dementia in Primary Care

Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Canadian Geriatrics Journal. 2012;15(3):85.
**187.** Sarkar U, Bonacum D, Strull W, et al. Challenges of making a diagnosis in the outpatient setting: a multi-site survey of primary care physicians. BMJ Quality & Safety. 2012;21(8):641-648.
**188.** Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. International Journal of Evidence-Based Healthcare. 2015;13(3):132-140.
**189.** Hartling L, Chisholm A, Thomson D, Dryden DM. A Descriptive Analysis of Overviews of Reviews Published between 2000 and 2011. PLOS ONE. 2012;7(11):e49667.
**190.** Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. Journal of Clinical Epidemiology.65(12):1267-1273.
**191.** Becker L, Oxman A. Chapter 22: Overviews of reviews. In: JPT H, S G. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 The Cochrane Collaboration; 2011.
**192.** Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. PLoS medicine. 2016;13(5):e1002028.
**193.** Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. The Milbank Quarterly. 2016;94(3):485-514.
**194.** Pollock M, Fernandes R, Becker L, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. Systematic Reviews. 2016;5.
**195.** Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC medical research methodology. 2011;11(1):15.
**196.** Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: A comparison of cochrane reviews with articles published in paper-based journals. JAMA. 1998;280(3):278-280.
**197.** Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. BMJ. 2006;333(7572):782.
**198.** Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting. The Cochrane Library. 2014.
**199.** Seitz DP, Fage BA, Chan CC, et al. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a primary care setting. The Cochrane Library. 2014.

**200.** Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. The Cochrane Library. 2015.

**201.** Hendry K, Lees RA, McShane R, Noel-Storr AH, Stott DJ, Quinn TJ. AD-8 for diagnosis of dementia across a variety of healthcare settings. The Cochrane Library. 2014.

**202.** Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. The Cochrane Library. 2014.

**203.** Fage BA, Chan CC, Gill SS, et al. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting. The Cochrane Library. 2015.

**204.** Arevalo-Rodriguez I, Smailagic N, Roqué i Figuls M, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). The Cochrane Library. 2015.

**205.** Creavin ST, Wisniewski S, Noel-Storr AH, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. The Cochrane Library. 2016.

**206.** Davis DH, Creavin ST, Yip JL, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. The Cochrane Library. 2015.

**207.** Lees R, Selvarajah J, Fenton C, et al. Test accuracy of direct to patient cognitive screening tests for diagnosis of post stroke cognitive impairment and dementia–Systematic review and meta-analysis. International Journal of Stroke. 2014;9:43.

**208.** Yokomizo JE, Simon SS, de Campos Bottino CM. Cognitive screening for dementia in primary care: a systematic review. International psychogeriatrics. 2014;26(11):1783-1804.

**209.** RCP. OP86: Individual patient outcome measures recommended for use in older people's mental health. http://www.rcpsych.ac.uk/usefulresources/publications/collegereports/op/op86.aspx 2012.

**210.** Ballard C, Burns A, Corbett A, Livingston G, Rasmussen J. Helping you to assess cognition: A practical toolkit for clinicians. England: Alzheimer s Society. 2013.

**211.** Macaskill P, Gatsonis C, Deeks JJ HR, Y. T. Chapter 10: Analysing and Presenting Results. In: Deeks JJ BP, Gatsonis C. Cochrane Handbook of Systematic Reviews of Diagnostic Test Accuracy Vol 1.0: The Cochrane Collaboratio; 2010.

**212.** Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology. 2007;7(1):10.

**213.** Whiting P, Savović J, Higgins JP, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. Journal of clinical epidemiology. 2016;69:225-234.

**214.** Clarke M, Jagger C, Anderson J, Battcock T, Kelly F, Stern MC. The prevalence of dementia in a total population: a comparison of two screening instruments. Age and Ageing. 1991;20(6):396-403.

**215.** Matthews FE, Arthur A, Barnes LE, et al. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. The Lancet. 2013;382(9902):1405-1412.

**216.** WIND AW, SCHELLEVIS FG, VAN STAVEREN G, SCHOLTEN RJ, JONKER C, VAN EIJK JTM. Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice. International journal of geriatric psychiatry. 1997;12(1):101-108.

**217.** Carnero-Pardo C, Espejo-Martínez B, López-Alcalde S, Espinosa García M, Feria Vilar I, L MN. ¿Es hora de jubilar al Mini-Mental? . Neurologia. 2008;23:648-649.

**218.** Belle SH, Mendelsohn AB, Seaberg EC, Ratcliff G. A brief cognitive screening battery for dementia in the community. Neuroepidemiology. 2000;19(1):43-50.
**219.** Wilder D, Cross P, Chen J, et al. Operating characteristics of brief screens for dementia in a multicultural population. The American Journal of Geriatric Psychiatry. 1995;3(2):96-107.
**220.** Carnero Pardo C, Espejo Martínez B, Montoro Rios M. Revisión sisteámtica y metaanálisis de la utilidad diagnóstica del Eurotest en la identificación de la demencia. Alzheimer. 2009(42):14-22.
**221.** Carnero-Pardo C, Lopez-Alcalde S, Allegri RF, Russo MJ. A systematic review and meta-analysis of the diagnostic accuracy of the Phototest for cognitive impairment and dementia. Dementia & Neuropsychologia. 2014;8(2):141-147.
**222.** Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. Journal of psychiatric research. 2009;43(4):411-431.
**223.** Kirby M, Denihan A, Bruce I, Coakley D, Lawlor BA. The clock drawing test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly. International Journal of Geriatric Psychiatry. 2001;16(10):935-940.
**224.** Borson S, Scanlan JM, Chen P, Ganguli M. The Mini-Cog as a screen for dementia: validation in a population-based sample. Journal of the American Geriatrics Society. 2003;51(10):1451-1454.
**225.** Kilada S, Gamaldo A, Grant EA, Moghekar A, Morris JC, O'brien RJ. Brief screening tests for the diagnosis of dementia: comparison with the mini-mental state exam. Alzheimer Disease & Associated Disorders. 2005;19(1):8-16.
**226.** Shaaban J, Aziz AA, Abdullah Z, Ab Razak A. Validation of the Malay Version of Rowland Universal Dementia Assessment Scale (MRUDAS) among Elderly Attending Primary Care Clinic. International Medical Journal. 2013;20(5).
**227.** Xu G, Meyer JS, Thornby J, Chowdhury M, Quach M. Screening for mild cognitive impairment (MCI) utilizing combined mini-mental-cognitive capacity examinations for identifying dementia prodromes. International journal of geriatric psychiatry. 2002;17(11):1027-1033.
**228.** Solomon PR, Pendlebury WW. Recognition of Alzheimer's disease: the 7 Minute Screen. Fam Med. 1998;30(4):265-271.
**229.** Lavery LL, Lu S-y, Chang C-CH, Saxton J, Ganguli M. Cognitive assessment of older primary care patients with and without memory complaints. Journal of general internal medicine. 2007;22(7):949-954.
**230.** Grober E, Hall C, Lipton RB, Teresi JA. Primary care screen for early dementia. Journal of the American Geriatrics Society. 2008;56(2):206-213.
**231.** Mitchell AJ, Malladi S. Screening and case-finding tools for the detection of dementia. Part II: evidence-based meta-analysis of single-domain tests. The American Journal of Geriatric Psychiatry. 2010;18(9):783-800.
**232.** Cruz-Orduña I, Bellón JM, Torrero P, et al. Detecting MCI and dementia in primary care: efficiency of the MMS, the FAQ and the IQCODE. Family practice. 2011;29(4):401-406.
**233.** Carnero-Pardo C, Espejo-Martinez B, Lopez-Alcalde S, et al. Effectiveness and costs of phototest in dementia and cognitive impairment screening. BMC neurology. 2011;11(1):92.
**234.** Modrego PJ, Gazulla J. The predictive value of the memory impairment screen in patients with subjective memory complaints: a prospective study. The primary care companion for CNS disorders. 2013;15(1).
**235.** Modrego PJ, Fayed N, Pina MA. Conversion from mild cognitive impairment to probable Alzheimer's disease predicted by brain magnetic resonance spectroscopy. American Journal of Psychiatry. 2005;162(4):667-675.
**236.** Brodaty H, Low L-F, Gibson L, Burns K. What is the best dementia screening instrument for general practitioners to use? The American journal of geriatric psychiatry. 2006;14(5):391-400.

**237.** Nakata E, Kasai M, Kasuya M, et al. Combined memory and executive function tests can screen mild cognitive impairment and converters to dementia in a community: the Osaki-Tajiri project. Neuroepidemiology. 2009;33(2):103-110.
**238.** Meguro K, Ishii H, Kasuya M, et al. Incidence of dementia and associated risk factors in Japan: The Osaki-Tajiri Project. Journal of the neurological sciences. 2007;260(1):175-182.
**239.** O'connor D, Pollitt P, Hyde J, et al. The reliability and validity of the Mini-Mental State in a British community survey. Journal of psychiatric research. 1989;23(1):87-96.
**240.** Brayne C, Calloway P. An epidemiological study of dementia in a rural population of elderly women. The British Journal of Psychiatry. 1989;155(2):214-219.
**241.** WIND AW, SCHELLEVIS FG, VAN STAVEREN GERRIT, SCHOLTEN RJ, JONKER C, VAN EIJK JTM. Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice. International journal of geriatric psychiatry. 1997;12(1):101-108.
**242.** Carnero-Pardo C, Sáez-Zea C, Montiel-Navarro L, Feria-Vilar I, Gurpegui M. Normative and reliability study of fototest. Neurología (English Edition). 2011;26(1):20-25.
**243.** Lourenço RA, Veras RP. Mini-Mental State Examination: psychometric characteristics in elderly outpatients. Revista de Saúde Pública. 2006;40(4):712-719.
**244.** Heun R, Papassotiropoulos A, Jennssen F. The validity of psychometric instruments for detection of dementia in the elderly general population. International journal of geriatric psychiatry. 1998;13(6):368-380.
**245.** Espejo-Martinez B, Carnero-Pardo C, Montoro-Ríos MT. Systematic review and meta-analysis of the accuracy of the eurotest for the detection of dementia. Alzheimer's & Dementia. 2009;5(4):P447.
**246.** Roth M, Tym E, Mountjoy CQ, et al. CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. The British journal of psychiatry : the journal of mental science. 1986;149:698-709.
**247.** Figueiredo S. Stroke Engine. 2009; Home/Assessments/Cognition/Cambridge Cognition Examination (CAMCOG). Available at.
**248.** Burns A, Lawlor B, Craig S. Assessment scales in old age psychiatry: Taylor & Francis; 2004.
**249.** POND CD, MANT A, KEHOE L, HEWITT H, BRODATY H. General practitioner diagnosis of depression and dementia in the elderly: can academic detailing make a difference? Family Practice. 1994;11(2):141-147.
**250.** Cullen B, Fahy S, Cunningham CJ, et al. Screening for dementia in an Irish community sample using MMSE: a comparison of norm-adjusted versus fixed cut-points. International journal of geriatric psychiatry. 2005;20(4):371-376.
**251.** Brodaty H, Kemp NM, Low LF. Characteristics of the GPCOG, a screening tool for cognitive impairment. International journal of geriatric psychiatry. 2004;19(9):870-874.
**252.** Carnero-Pardo C, Saez-Zea C, De la Vega Cotarelo R, Gurpegui M. FOTOTRANS Study. Multicentre study on the validity of Fototest under clinical practice conditions. Neurología (English Edition). 2012;27(2):68-75.
**253.** Pardo CC, de la Vega Cotarelo R, Alcalde SL, et al. Assessing the diagnostic accuracy (DA) of the Spanish version of the informant-based AD8 questionnaire. Neurología (English Edition). 2013;28(2):88-94.
**254.** Borson S, Scanlan JM, Watanabe J, Tu SP, Lessig M. Simplifying detection of cognitive impairment: comparison of the Mini-Cog and Mini-Mental State Examination in a multiethnic sample. Journal of the American Geriatrics Society. 2005;53(5):871-874.
**255.** Carnero-Pardo C, Cruz-Orduña I, Espejo-Martínez B, Martos-Aparicio C, López-Alcalde S, Olazarán J. Utility of the Mini-Cog for detection of cognitive impairment in

Primary Care: data from two spanish studies. International Journal of Alzheimer's Disease. 2013;2013.
**256.** Holsinger T, Plassman BL, Stechuchak KM, Burke JR, Coffman CJ, Williams JW. Screening for cognitive impairment: comparing the performance of four instruments in primary care. Journal of the American Geriatrics Society. 2012;60(6):1027-1036.
**257.** Fuchs A, Wiese B, Altiner A, Wollny A, Pentzek M. Cued recall and other cognitive tasks to facilitate dementia recognition in primary care. Journal of the American Geriatrics Society. 2012;60(1):130-135.
**258.** Grober E, Hall C, Mcginn M, et al. Neuropsychological strategies for detecting early dementia. Journal of the International Neuropsychological Society. 2008;14(1):130-142.
**259.** Lipton RB, Katz MJ, Kuslansky G, et al. Screening for dementia by telephone using the memory impairment screen. Journal of the American Geriatrics Society. 2003;51(10):1382-1390.
**260.** Kuslansky G, Buschke H, Katz M, Sliwinski M, Lipton RB. Screening for Alzheimer's disease: the memory impairment screen versus the conventional three-word memory test. Journal of the American Geriatrics Society. 2002;50(6):1086-1091.
**261.** Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting. The Cochrane database of systematic reviews. 2014;7:CD010771.
**262.** Cochrane. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 In: Higgins JPT, S G, eds. Available from http://handbook.cochrane.org.: The Cochrane Collaboration, 2011. ; 2011.
**263.** Cochrane. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Available from: http://srdta.cochrane.org/. The Cochrane Collaboration; 2013.
**264.** Naugle RI, Kawczak K. Limitations of the Mini-Mental State Examination. Cleve Clin J Med. 1989;56(3):277-281.
**265.** Nieuwenhuis-Mark E. The death knoll for the MMSE: has it outlived its purpose? Journal of Geriatric Psychiatry and Neurology 23 (3), 151-7. 2010.
**266.** Brown J. The use and misuse of short cognitive tests in the diagnosis of dementia. J Neurol Neurosurg Psychiatry. 2015;86(6):680-685.
**267.** Wolf-Klein GP, Silverstone FA, Levy AP, Brod MS, Breuer J. Screening for Alzheimer's disease by clock drawing. Journal of the American Geriatrics Society. 1989;37(8):730-734.
**268.** Ball LJ, Ogden A, Mandi D, Birge SJ. The validation of a mailed health survey for screening of dementia of the Alzheimer's type. Journal of the American Geriatrics Society. 2001;49(6):798-802.
**269.** Del Ser T, Sánchez-Sánchez F, de Yébenes MJG, Otero A, Munoz DG. Validation of the seven-minute screen neurocognitive battery for the diagnosis of dementia in a Spanish population-based sample. Dementia and geriatric cognitive disorders. 2006;22(5-6):454-464.
**270.** Shulman KI, Shedletsky R, Silver IL. The challenge of time: clock-drawing and cognitive function in the elderly. International journal of geriatric psychiatry. 1986;1(2):135-140.
**271.** Sunderland T, Hill JL, Mellow AM, et al. Clock drawing in Alzheimer's disease. Journal of the American Geriatrics Society. 1989;37(8):725-729.
**272.** McKenzie JE, Brennan SE. Overviews of systematic reviews: great promise, greater challenge: BioMed Central; 2017.
**273.** Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview Systematic Reviews2017.
**274.** Pollock M, Fernandes RM, Becker LA, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare

interventions? A scoping review and qualitative metasummary. Systematic reviews. 2016;5(1):190.

**275.** Pollock A, Campbell P, Brunton G, Hunt H, Estcourt L. Selecting and implementing overview methods: implications from five exemplar overviews. Systematic reviews. 2017;6(1):145.

**276.** Pollock M, Fernandes R, Becker L, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. Systematic Reviews. 2016 5(1).

**277.** Worswick J, Wayne SC, Bennett R, et al. Improving quality of care for persons with diabetes: an overview of systematic reviews - what does the evidence tell us? Systematic Reviews. 2013;2(1):26.

**278.** Carnero Pardo C, Espejo Martínez B, Montoro Rios MT. Revisión sisteámtica y metaanálisis de la utilidad diagnóstica del Eurotest en la identificación de la demencia. [Systematic review and meta-analysis of the diagnostic utility of Eurotest in identifying the dementia]. Alzheimer. 2009;42: 14-22.

**279.** Fage BA, Chan CCH, Gill SS, et al. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting. Cochrane Database of Systematic Reviews. 2015(2).

**280.** Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. The Cochrane database of systematic reviews. 2015(3):Cd010772.

**281.** Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. Cochrane Database of Systematic Reviews. 2014(4).

**282.** Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Chapter 11: Interpreting results and drawing conclusions. Cochrane handbook for systematic reviews of diagnostic test accuracy version 0.9. The Cochrane Collaboration. 2013.

**283.** A review of existing systematic reviews summarising the accuracy of brief cognitive assessments for identifying dementia, particularly for use in primary care. 2015. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015022078.

**284.** Hunt HA, Van Kampen S, Takwoingi Y, Llewellyn DJ, Pearson M, Hyde CJ. The comparative diagnostic accuracy of the Mini Mental State Examination (MMSE) and the General Practitioner assessment of Cognition (GPCOG) for identifying dementia in primary care: a systematic review protocol. Diagnostic and Prognostic Research. 2017;1(1):14.

**285.** Harrison JK, Fearon P, Noel-Storr A, McShane R, Stott D, Quinn T. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting. Cochrane Database of Systematic Reviews. 2014(7).

**286.** Pirani A, Brodaty H, Martini E, Zaccherini D, Neviani F, Neri M. The validation of the Italian version of the GPCOG (GPCOG-It): a contribution to cross-national implementation of a screening test for dementia in general practice. International Psychogeriatrics. 2010;22(01):82-90.

**287.** Basic D, Khoo A, Conforti D, et al. Rowland Universal Dementia Assessment Scale, Mini-Mental State Examination and General Practitioner Assessment of Cognition in a multicultural cohort of community-dwelling older persons with early dementia. Australian Psychologist. 2009;44(1):40-53.

**288.** Li X, Xiao S, Fang Y, et al. Validation of the general practitioner assessment of cognition–Chinese version (GPCOG-C) in China. International psychogeriatrics. 2013;25(10):1649-1657.

**289.** Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Systematic reviews. 2016;5(1):210.

**290.** Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The Mini-Cog: a cognitive'vital signs' measure for dementia screening in multi-lingual elderly. International journal of geriatric psychiatry. 2000;15(11):1021-1027.
**291.** Nilsson FM. Mini Mental State Examination (MMSE) – probably one of the most cited papers in health science. Acta Psychiatrica Scandinavica. 2007;116(2):156-157.
**292.** de Koning I, van Kooten F, Dippel DW, et al. The CAMCOG: a useful screening instrument for dementia in stroke patients. Stroke. 1998;29(10):2080-2086.
**293.** Canadian Partnership for Stroke Recovery. Stroke Engine: Cambridge Cognition Examination (CAMCOG). 2017; Home » Assessments » Cognition » Cambridge Cognition Examination (CAMCOG)

Description of the CAMCOG instrument. Available at: https://www.strokengine.ca/indepth/camcog_indepth/. Accessed 11/09/2017, 2017.
**294.** Martin R, O'Neill D. Taxing your memory. The Lancet. 2009;373(9680).
**295.** Newman JC, Feldman R. Copyright and open access at the bedside. New England Journal of Medicine. 2011;365(26):2447-2449.
**296.** Carnero-Pardo C. Should the mini-mental state examination be retired? Neurología (English Edition). 2014;29(8):473-481.
**297.** APA. Diagnostic and Statistical Manual of Mental Disorders (DSM–5). 2017; http://www.dsm5.org/psychiatrists/practice/dsm. Accessed 27/01/2017, 2017.
**298.** WHO. The 11th Revision of the International Classification of Diseases (ICD-11) is here. 2018; http://www.who.int/classifications/icd/en/. Accessed 29/09/2018, 2018.
**299.** Whiting PF, Rutjes AS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine. 2011;155(8):529-536.
**300.** P M, C G, JJ D, RM H, Y. T. Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*2010: http://srdta.cochrane.org/.
**301.** Rifai N, Altman DG, Bossuyt PM. Reporting bias in diagnostic and prognostic studies: time for action: Clinical Chemistry; 2008.
**302.** Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case–control and two-gate designs in diagnostic accuracy studies: Clinical Chemistry; 2005.
**303.** BURNS A, LAWLOR B, CRAIG S. Rating scales in old age psychiatry. The British Journal of Psychiatry. 2002;180(2):161-167.
**304.** Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clinical Chemistry. 2008;54(4):729-737.
**305.** Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Annals of internal medicine. 2004;140(3):189-202.
**306.** Velayudhan L, Ryu S-H, Raczek M, et al. Review of brief cognitive tests for patients with suspected dementia. International Psychogeriatrics / Ipa. 2014;26(8):1247-1262.
**307.** Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology. 2010;5(9):1315-1316.
**308.** Lijmer JG, Leeflang M, Bossuyt PM. Proposals for a phased evaluation of medical tests. Medical Decision Making. 2009;29(5):E13-E21.
**309.** Replogle WH, Johnson WD, Hoover KW. Using evidence to determine diagnostic test efficacy. Worldviews on evidence-based nursing. 2009;6(2):87-92.
**310.** Jedenius E, Wimo A, Stromqvist J, Jonsson L, Andreasen N. The cost of diagnosing dementia in a community setting. Int J Geriatr Psychiatry. 2010;25(5):476-482.
**311.** Wimo A, Religa D, Spångberg K, Edlund AK, Winblad B, Eriksdotter M. Costs of diagnosing dementia: results from SveDem, the Swedish Dementia Registry. International journal of geriatric psychiatry. 2013;28(10):1039-1044.

**312.** England N. NHS England, Government and BMA agree new GP contract for 2016/17. 2016; https://www.england.nhs.uk/2016/02/gp-contract-16-17/. Accessed 27/02/2016, 2016.
**313.** Bossuyt PM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Medical Decision Making. 2009;29(5):E30-E38.
**314.** Dhedhi SA, Swinglehurst D, Russell J. 'Timely'diagnosis of dementia: what does it mean? A narrative analysis of GPs' accounts. BMJ open. 2014;4(3):e004439.
**315.** Carr DB, O'Neill D. Mobility and safety issues in drivers with dementia. International psychogeriatrics. 2015;27(10):1613-1622.
**316.** Adler G, Rottunda SJ. Ethical Considerations for the Driver with Dementia. Ethical Considerations and Challenges in Geriatrics: Springer; 2017.
**317.** Connors MH, Ames D, Woodward M, Brodaty H. Predictors of driving cessation in dementia: baseline characteristics and trajectories of disease progression. Alzheimer Disease & Associated Disorders. 2018;32(1):57-61.
**318.** Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine. 2013;20(11):1194-1206.
**319.** Wade R, Corbett M, Eastwood A. Quality assessment of comparative diagnostic accuracy studies: our experience using a modified version of the QUADAS-2 tool. Research synthesis methods. 2013;4(3):280-286.
**320.** Willis BH. Spectrum bias—why clinicians need to be cautious when applying diagnostic test studies. Family practice. 2008;25(5):390-396.
**321.** Pentzek M, Vollmar HC, Wilm S, Leve V. Putting dementia awareness into general practice. Zeitschrift für Gerontologie und Geriatrie. 2017;50(2):44-47.
**322.** Mate KE, Magin PJ, Brodaty H, et al. An evaluation of the additional benefit of population screening for dementia beyond a passive case-finding approach. International Journal of Geriatric Psychiatry. 2017;32(3):316-323.
**323.** Graham J. New Toolkits Help Physicians Detect, Diagnose, and Manage Dementia. Jama. 2017;318(14):1310-1312.
**324.** Pond CD, Mate KE, Phillips J, et al. Predictors of agreement between general practitioner detection of dementia and the revised Cambridge Cognitive Assessment (CAMCOG-R). International psychogeriatrics. 2013;25(10):1639-1647.
**325.** GSA. The Gerontological Society of America Workgroup on Cognitive Impairment Detection and Earlier Diagnosis: Report and recommendations 2015.
**326.** Health Do. Prime Minister's Challenge on Dementia 2020 2016.
**327.** Committee GA. Clinical Practice Guidelines and Principles of Care for People with Dementia

Sydney: Guideline Adaptation Committee; 2016.
**328.** Moore V, Cahill S. Diagnosis and disclosure of dementia–A comparative qualitative study of Irish and Swedish General Practitioners. Aging & Mental Health. 2013;17(1):77-84.
**329.** van Hout HP, Vernooij-Dassen M, Bakker K, Blom M, Grol R. General practitioners on dementia: tasks, practices and obstacles. Patient Educ Couns. 2000;39.
**330.** Ólafsdóttir M, Foldevi M, Marcusson J. Dementia in primary care: why the low detection rate? Scandinavian journal of primary health care. 2001;19(3):194-198.
**331.** Turner S, Iliffe S, Downs M, et al. General practitioners' knowledge, confidence and attitudes in the diagnosis and management of dementia. Age Ageing. 2004;33.
**332.** Tang EYH, Birdi R, Robinson L. Attitudes to diagnosis and management in dementia care: views of future general practitioners. International Psychogeriatrics. 2016:1-6.
**333.** Milne AJ, Hamilton-West K, Hatzidimitriadou E. GP attitudes to early diagnosis of dementia: evidence of improvement. Aging & Mental Health. 2005;9(5):449-455.
**334.** Ahmad S, Orrell M, Iliffe S, Gracie A. GPs' attitudes, awareness, and practice regarding early diagnosis of dementia. Br J Gen Pract. 2010;60(578):e360-e365.

**335.** Renshaw J, Scurfield P, Cloke L, Orrell M. General practitioners' views on the early diagnosis of dementia. Br J Gen Pract. 2001;51(462):37-38.
**336.** ADI. World Alzheimer Report 2016: Improving healthcare for people living with dementia. Coverage, Quality and costs now and in the future 2016.
**337.** Health Do. G8 Dementia Summit: global action against dementia 2013.
**338.** Pachana NA, Mitchell LK, Pinsker DM, Morriss E, Lo A, Cherrier M. In Brief, Look Sharp: Short Form Assessment in the Geriatric Setting. Australian Psychologist. 2016;51(5):342-351.
**339.** Iracleous P, Nie JX, Tracy CS, et al. Primary care physicians' attitudes towards cognitive screening: findings from a national postal survey. International journal of geriatric psychiatry. 2010;25(1):23-29.
**340.** Foley T, Boyle S, Jennings A, Smithson WH. "We're certainly not in our comfort zone": a qualitative study of GPs' dementia-care educational needs. BMC Family Practice. 2017;18(1):66.
**341.** Boise L, Camicioli R, Morgan DL, Rose JH, Congleton L. Diagnosing dementia: perspectives of primary care physicians. Gerontologist. 1999;39.
**342.** Downs M, Cook A, Rae C, Collins KE. Caring for patients with dementia: the GP perspective. Aging Mental Health. 2000;4.
**343.** Shulman KI, Herrmann N, Brodaty H, et al. IPA survey of brief cognitive screening instruments. International Psychogeriatrics. 2006;18(2):281-294.
**344.** Gibbs S. Passwords and hacking: the jargon of hashing, salting and SHA-2 explained. The Guardian2016: Data and computer security.
**345.** Iliffe S, Robinson L, Brayne C, et al. Primary care and dementia: 1. diagnosis, screening and disclosure. International journal of geriatric psychiatry. 2009;24(9):895-901.
**346.** Hansen EC, Hughes C, Routley G, Robinson AL. General practitioners' experiences and understandings of diagnosing dementia: Factors impacting on early diagnosis. Social science & medicine. 2008;67(11):1776-1783.
**347.** Dodd E, Cheston R, Cullum S, et al. Primary care-led dementia diagnosis services in South Gloucestershire: Themes from people and families living with dementia and health care professionals. Dementia. 2016;15(6):1586-1604.
**348.** Milne AJ, Woolford H, Mason J, Hatzidimitriadou E. Early diagnosis of dementia by GPs: an exploratory study of attitudes. Aging & Mental Health. 2000;4(4):292-300.
**349.** Wijeratne C, Harris P. Late life depression and dementia: a mental health literacy survey of Australian general practitioners. Int Psychogeriatr. 2009;21.
**350.** Teel CS. Rural practitioners' experiences in dementia diagnosis and treatment. Aging Mental Health. 2004;8.
**351.** Pathak KP, Montgomery A. General practitioners' knowledge, practices, and obstacles in the diagnosis and management of dementia. Aging & Mental Health. 2015;19(10):912-920.
**352.** Murphy K, O'Connor DA, Browning CJ, et al. Understanding diagnosis and management of dementia and guideline implementation in general practice: a qualitative study using the theoretical domains framework. Implementation Science. 2014;9(1):31.
**353.** Gove D, Small N, Downs M, Vernooij-Dassen M. General practitioners' perceptions of the stigma of dementia and the role of reciprocity. Dementia. 2016:1471301215625657.
**354.** Hardwick R, Heaton J, Griffiths G, et al. Exploring reasons for variation in ordering thyroid function tests in primary care: a qualitative study. Quality in Primary Care. 2014;22(6):256-261.
**355.** Davenport C. Systematic reviews and meta-analyses of test accuracy: developing methods for conducting and presenting results of systematic reviews of test accuracy. etheses.bham.ac.uk: University of Birmingham; 2012.

**356.** Chisnell J, Zhelev Z, Hyde CJ, Hamilton W. Improving Cancer Referral Decision Support in Primary Care: Results of a GP Survey (unpublished draft)2017. Located at: n/a, n/a.
**357.** Mason J. Qualitative researching: Sage; 2002.
**358.** Robson C, McCartan K. Real world research: John Wiley & Sons; 2016.
**359.** Attride-Stirling J. Thematic networks: an analytic tool for qualitative research. Qualitative research. 2001;1(3):385-405.
**360.** Gonçalves-Bradley DC, Boylan A-M, Koshiaris C, Montes MV, Ford GA, Lasserson DS. GPs' adherence to guidelines for structured assessments of stroke survivors in the community and care homes. Family practice. 2015;32(6):659-663.
**361.** Garcia J, Evans J, Reshaw M. ``Is There Anything Else You Would Like to Tell Us''–Methodological Issues in the Use of Free-Text Comments from Postal Surveys. Quality & Quantity. 2004;38(2):113-125.
**362.** Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). Journal of medical Internet research. 2004;6(3).
**363.** Carvalho S, White H. Combining the quantitative and qualitative approaches to poverty measurement and analysis: the practice and the potential: The World Bank; 1997.
**364.** Cohen L, Manion L, Morrison K. Research methods in education: routledge; 2002.
**365.** Rosen CS, Chow HC, Greenbaum MA, et al. How well are clinicians following dementia practice guidelines? Alzheimer Disease & Associated Disorders. 2002;16(1):15-23.
**366.** Wilcock J, Iliffe S, Turner S, et al. Concordance with clinical practice guidelines for dementia in general practice. Aging and Mental Health. 2009;13(2):155-161.
**367.** Downs M, Turner S, Bryans M, et al. Effectiveness of educational interventions in improving detection and management of dementia in primary care: cluster randomised controlled study. Bmj. 2006;332(7543):692-696.
**368.** Millard FB, Thistlethwaite J, Baune BT, Kennedy RL, Spagnolo C. Dementia Diagnosis: A Pilot Randomised Controlled Trial of Education and IT Audit to Assess Change in GP Dementia Documentation. Australian Journal of Primary Health. 2008;14(3):141-149.
**369.** Dooley J, Bailey C, McCabe R. Communication in healthcare interactions in dementia: a systematic review of observational studies. International Psychogeriatrics. 2015;27(8):1277-1300.
**370.** Wilcock J, Jain P, Griffin M, et al. Diagnosis and management of dementia in family practice. Aging & Mental Health. 2016;20(4):362-369.
**371.** NICE. clinical guideline 42. Dementia: Supporting people with dementia and their carers in health and social care. In: Health Do2006.
**372.** Health Do. Living well with dementia: A national dementia strategy: Department of Health; 2009.
**373.** Orton PK, Gray DP. Factors influencing consultation length in general/family practice. Family Practice. 2016;33(5):529-534.
**374.** WIND AW, SCHELLEVIS FG, VAN STAVEREN GERRIT, SCHOLTEN RJ, JONKER C, & , VAN EIJK JTM. Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice. International journal of geriatric psychiatry. 1997;12(1):101-108.
**375.** Milne A. Dementia screening and early diagnosis: The case for and against. Health, Risk & Society. 2010;12(1):65-76.
**376.** Parmar J, Dobbs B, McKay R, et al. Diagnosis and management of dementia in primary care: exploratory study. Canadian Family Physician. 2014;60(5):457-465.
**377.** Black BS, Johnston D, Rabins PV, Morrison A, Lyketsos C, Samus QM. Unmet needs of community-residing persons with dementia and their informal caregivers: Findings from the maximizing independence at home study. Journal of the American Geriatrics Society. 2013;61(12):2087-2095.

**378.** Schoenmakers B, Buntinx F, Delepeleire J. What is the role of the general practitioner towards the family caregiver of a community-dwelling demented relative? A systematic literature review. Scandinavian journal of primary health care. 2009;27(1):31-40.
**379.** Iliffe S, Robinson, L., Brayne C, Goodman, C., Rait G, Manthorpe J, Ashley P. Primary care and dementia: 1. diagnosis, screening and disclosure. Int. J. Geriat. Psychiatry. 2009;24:895–901.
**380.** Manthorpe J, Samsi K, Campbell S, et al. From forgetfulness to dementia: clinical and commissioning implications of diagnostic experiences. Br J Gen Pract. 2013;63(606):e69-e75.
**381.** Litchfield I, Bentham L, Hill A, McManus RJ, Lilford R, Greenfield S. Routine failures in the process for blood testing and the communication of results to patients in primary care in the UK: a qualitative exploration of patient and provider perspectives. BMJ Qual Saf. 2015;24(11):681-690.
**382.** Nieuwenhuis-Mark E. The death knoll for the MMSE: has it outlived its purpose? . Journal of Geriatric Psychiatry and Neurology 23 (3), 151-7. (2010)
**383.** NICE. Dementia: assessment, management and support for people living with dementia and their carers. In: Excellence NIfHaC. Vol NG 97. online: NICE; 2018.
**384.** Pink J, O'Brien J, Robinson L, Longson D. Dementia: assessment, management and support: summary of updated NICE guidance. BMJ. 2018;361.
**385.** Apolinario D, Lichtenthaler DG, Magaldi RM, et al. Using temporal orientation, category fluency, and word recall for detecting cognitive impairment: the 10-point cognitive screener (10-CS). International journal of geriatric psychiatry. 2016;31(1):4-12.
**386.** Abdel-Aziz K, Larner A. Six-item cognitive impairment test (6CIT): pragmatic diagnostic accuracy study for dementia and MCI. International psychogeriatrics. 2015;27(6):991-997.
**387.** Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. Medical care. 2002;40(9):771-781.
**388.** Hancock P, Larner A. Test Your Memory test: diagnostic utility in a memory clinic population. International journal of geriatric psychiatry. 2011;26(9):976-980.
**389.** Postel-Vinay N, Hanon O, Clerson P, et al. Validation of the Test Your Memory (F-TYM test) in a French memory clinic population. The Clinical Neuropsychologist. 2014;28(6):994-1007.
**390.** Petrazzuoli F, Vinker S, Koskela TH, et al. Exploring dementia management attitudes in primary care: a key informant survey to primary care physicians in 25 European countries. International Psychogeriatrics. 2017;29(9):1413-1423.
**391.** Sarkar U, Bonacum D, Strull W, et al. Challenges of making a diagnosis in the outpatient setting: a multi-site survey of primary care physicians. BMJ Quality & Safety. 2012;21(8):641-648.
**392.** Chan JY, Kwong JS, Wong A, Kwok TC, Tsoi KK. Comparison of Computerized and Paper-and-Pencil Memory Tests in Detection of Mild Cognitive Impairment and Dementia: A Systematic Review and Meta-analysis of Diagnostic Studies. Journal of the American Medical Directors Association. 2018.
**393.** Bruun M, Rhodius-Meester HF, Koikkalainen J, et al. Evaluating combinations of diagnostic tests to discriminate different dementia types. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2018;10:509-518.
**394.** McCleery J, Flicker L, Richard E, Quinn TJ. When is Alzheimer's not dementia—Cochrane commentary on The National Institute on Ageing and Alzheimer's Association Research Framework for Alzheimer's Disease. Age and Ageing. 2018:afy167-afy167.
**395.** Menon R, Larner AJ. Use of cognitive screening instruments in primary care: the impact of national dementia directives (NICE/SCIE, National Dementia Strategy). Family Practice. 2011;28(3):272-276.

**396.** Health NCCfM. Dementia: supporting people with Dementia and their carers in health and social care (NICE Clinical Guideline 42). National Institute for Health and Clinical Excellence. 2006.
**397.** Banerjee S, Owen J. Living well with dementia: a national dementia strategy. London: Department of Health. 2009.
**398.** Petrovitch H, White LR, Ross GW, et al. Accuracy of clinical criteria for AD in the Honolulu–Asia Aging Study, a population-based study. Neurology. 2001;57(2):226-234.
**399.** Kazee A, Eskin T, Lapham L, Gabriel K, McDaniel K, Hamill R. Clinicopathologic correlates in Alzheimer disease: assessment of clinical and pathologic diagnostic criteria. Alzheimer disease and associated disorders. 1993.
**400.** Lim A, Tsuang D, Kukull W, et al. Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series. Journal of the American Geriatrics Society. 1999;47(5):564-569.
**401.** Zekry D, Hauw JJ, Gold G. Mixed dementia: epidemiology, diagnosis, and treatment. Journal of the American Geriatrics Society. 2002;50(8):1431-1438.
**402.** Aldus C, Arthur A, Fox C, et al. Cognitive function and ageing study ii dementia diagnosis study (caddy): The prevalence, causes and consequences of dementia undetected or undiagnosed in primary care in england. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2018;14(7):P573-P574.
**403.** International AsD. From Plan To Impact: Progress Of The Targets Towards Implementation Of The Global Action Plan. 2018.
**404.** Brooker D, Fontaine JL, Evans S, Bray J, Saad K. Public health guidance to facilitate timely diagnosis of dementia: ALzheimer's COoperative Valuation in Europe recommendations. 2014;29(7):682-693.
**405.** NHS. Dementia diagnosis and management: a brief pragmatic resource for general practitioners2015.
**406.** Organization WH. Global action plan on the public health response to dementia 2017–2025. 2017.
**407.** Laron M, Mannheim I, Brodsky J, et al. Barriers And Enablers To Timely Diagnosis Of Dementia: Patients'and Families' Points Of View. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2018;14(7):P1642.
**408.** Lilly. New International Survey Reveals Multiple Barriers to an Accurate and Timely Alzheimer's Disease Diagnosis. In: PRNewswire2012.
**409.** Fan W, Yen Z. Factors affecting response rates of the web survey: A systematic review. Comput Hum Behav. 2009;26.
**410.** Creavin ST, Noel-Storr AH, Richard E, et al. Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people. Cochrane Database of Systematic Reviews. 2017(2).
**411.** O'Connor D, Fertig A, Grande M, et al. Dementia in general practice: the practical consequences of a more positive approach to diagnosis. Br J Gen Pract. 1993;43(370):185-188.
**412.** Pentzek M, Fuchs A, Wiese B, et al. General practitioners' judgment of their elderly patients' cognitive status. Journal of general internal medicine. 2009;24(12):1314.
**413.** Bell S, Harkness K, Dickson J, Blackburn D. A diagnosis for £55: what is the cost of government initiatives in dementia case finding. Age and ageing. 2015;44(2):344-345.
**414.** Burns A. Alistair Burns and 51 colleagues reply to David Le Couteur and colleagues. . BMJ. 2013;347(oct15_6):f6125.
**415.** Rasmussen J. Would doctors routinely asking older patients about their memory improve dementia outcomes? Yes. Bmj. 2013;346:f1780.
**416.** McCartney M. Would doctors routinely asking older patients about their memory improve dementia outcomes? No. Bmj. 2013;346:f1745.
**417.** van den Dungen P, van Charante EPM, van de Ven PM, van Marwijk HW, van der Horst HE, van Hout HP. Case finding of mild cognitive impairment and dementia and

subsequent care; results of a cluster RCT in primary care. PloS one. 2016;11(6):e0156958.
**418.** Burn A-M, Fleming J, Brayne C, Fox C, Bunn F. Dementia case-finding in hospitals: a qualitative study exploring the views of healthcare professionals in English primary care and secondary care. BMJ open. 2018;8(3):e020521.
**419.** Force USPST. Final Recommendation Statement: Cognitive Impairment in Older Adults: Screening. . 2016; https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/cognitive-impairment-in-older-adults-screening.
**420.** Committee NS. Screening for dementia: External review against programme appraisal criteria for the UK National Screening Committee 2018.
**421.** McCartney M, Fell G, Muir Grey J, et al. Case Finding versus Screening. BMJ: Analysis.
**422.** Solomon PR, Pendlebury WW. Recognition of Alzheimer's disease: the 7 Minute Screen. . Fam Med. 1998;30(4):265-271.
**423.** Banerjee S, Smith SC, Lamping DL, et al. Quality of life in dementia: more than just cognition. An analysis of associations with quality of life in dementia. Journal of neurology, neurosurgery, and psychiatry. 2006;77(2):146-148.
**424.** Devine M. Patient centred diagnosis of dementia: we must listen to patients' wishes. BMJ: British Medical Journal (Online). 2017;359.
**425.** Fox C, Lafortune L, Boustani M, Brayne C. The pros and cons of early diagnosis in dementia. Br J Gen Pract. 2013.
**426.** Liu D, Green E, Kasteridis P, et al. The unintended consequences of primary care incentive schemes to increase dementia diagnoses in England. British Journal of General Practice. 2018.
**427.** Xanthopoulou P, Dooley J, Meo I, Bass N, McCabe R. Patient and companion concerns when receiving a dementia diagnosis: an observational study of dementia diagnosis feedback meetings. Ageing and Society. 2018:1-24.
**428.** Tong T, Thokala P, McMillan B, Ghosh R, Brazier J. Cost effectiveness of using cognitive screening tests for detecting dementia and mild cognitive impairment in primary care. Int J Geriatr Psychiatry. 2017;32(12):1392-1400.
**429.** Bruscoli M, Lovestone S. Is MCI really just early dementia? A systematic review of conversion studies. International Psychogeriatrics. 2004;16(2):129-140.
**430.** Olazarán J, Torrero P, Cruz I, et al. MCI and dementia in primary care: the value of medical history. Family practice. 2011:cmr005.
**431.** Chêne G, Favary C, Dubois B, et al. Short-term prediction of conversion from mild cognitive impairment to dementia in patients in the memento cohort, 2011-2016. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2017;13(7):P578-P579.
**432.** Richardson C, Stephan BCM, Robinson L, Matthews F. Risk of two-year progression from mild cognitive impairment to dementia: Results from the cognitive function and ageing study. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2018;14(7):P574.
**433.** Solfrizzi V, Panza F, Colacicco AM, et al. Vascular risk factors, incidence of MCI, and rates of progression to dementia. Neurology. 2004;63(10):1882-1891.
**434.** Tay L, Lim WS, Chan M, et al. New DSM-V neurocognitive disorders criteria and their impact on diagnostic classifications of mild cognitive impairment and dementia in a memory clinic setting. The American Journal of Geriatric Psychiatry. 2015;23(8):768-779.
**435.** Public Health England. National General Practice Profiles. 2016; This spine chart provides a summary of practice demography, deprivation, patient satisfaction and life expectancy estimates. Available at: http://fingertips.phe.org.uk/profile/general-practice/data#mod,1,pyr,2015,pat,19,par,-,are,-,sid1,2000005,ind1,-,sid2,-,ind2,-. Accessed 01/03/2016, 2016.

**436.** Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ. 2013;185.
**437.** Matthews FE, Arthur A, Barnes LE, et al. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. The Lancet. 2013;382(9902):1405-1412.
**438.** Basic D, Khoo A, Conforti D, et al. Rowland Universal Dementia Assessment Scale, Mini-Mental State Examination and General Practitioner Assessment of Cognition in a multicultural cohort of community-dwelling older persons with early dementia. Australian Psychologist. 2009;44(1):40-53.
**439.** Clague F, Mercer SW, McLean G, Reynish E, Guthrie B. Comorbidity and polypharmacy in people with dementia: insights from a large, population-based cross-sectional analysis of primary care data. Age and Ageing. 2017;46(1):33-39.
**440.** Bunn F, Burn AM, Goodman C, et al. Comorbidity and dementia: a mixed-method study on improving health care for people with dementia (CoDem). Health Services and Delivery Research. Southampton (UK): NIHR Journals Library



**441.** Bunn F, Burn A-M, Goodman C, et al. Comorbidity and dementia: a scoping review of the literature. BMC medicine. 2014;12(1):192.
**442.** Snowden MB, Atkins DC, Steinman LE, et al. Longitudinal association of dementia and depression. The American Journal of Geriatric Psychiatry. 2015;23(9):897-905.
**443.** Wu Y-T, Fratiglioni L, Matthews FE, et al. Dementia in western Europe: epidemiological evidence and implications for policy making. The Lancet Neurology. 2016;15(1):116-124.
**444.** Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One. 2007;2.
**445.** Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of clinical epidemiology. 2005;58(9):882-893.
**446.** Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. The Lancet. 2014;383(9913):267-276.
**447.** Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview. Systematic reviews. 2018;7(1):39.
**448.** Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. bmj. 2017;358:j4008.
**449.** Gates A, Gates M, Duarte G, et al. Evaluation of the reliability, usability, and applicability of AMSTAR, AMSTAR 2, and ROBIS: protocol for a descriptive analytic study. Systematic reviews. 2018;7(1):85.
**450.** van den Dungen P, van Marwijk HW, van der Horst HE, et al. The accuracy of family physicians' dementia diagnoses at different stages of dementia: a systematic review. International Journal of Geriatric Psychiatry. 2012;27(4):342-354.

**451.** di Ruffano LF, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. Bmj. 2012;344:e686.
**452.** Leeflang MMG, Bossuyt PMM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. Journal of Clinical Epidemiology. 2009;62(1):5-12.
**453.** Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. Annals of internal medicine. 2008;149(12):889-897.
**454.** Ochodo EA, Haan MCd, Reitsma JB, Hooft L, Bossuyt PM, Leeflang MMG. Overinterpretation and Misreporting of Diagnostic Accuracy Studies: Evidence of "Spin". Radiology. 2013;267(2):581-588.
**455.** Korevaar DA, Wang J, Enst WAv, et al. Reporting Diagnostic Accuracy Studies: Some Improvements after 10 Years of STARD. Radiology. 2015;274(3):781-789.
**456.** Noel-Storr AH, McCleery JM, Richard E, et al. Reporting standards for studies of diagnostic test accuracy in dementia: The STARDdem Initiative. Neurology. 2014;83(4):364-373.
**457.** Gopalakrishna G, Langendam MW, Scholten RJPM, Bossuyt PMM, Leeflang MMG. Defining the clinical pathway in cochrane diagnostic test accuracy reviews. BMC Medical Research Methodology. 2016;16(1):153.
**458.** Kitzinger J. Qualitative research: introducing focus groups. Bmj. 1995;311(7000):299-302.
**459.** Wong LP. Focus group discussion: a tool for health and medical research. Singapore Med J. 2008;49(3):256-260.
**460.** Sood R, Sood A, Ghosh AK. Non-evidence-based variables affecting physicians' test-ordering tendencies: a systematic review. Neth J Med. 2007;65(5):167-177.
**461.** Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. J Clin Epidemiol. 1997;50.
**462.** Beebe TJ, Locke GR, Barnes SA, Davern ME, Anderson KJ. Mixing web and mail methods in a survey of physicians. Health Serv Res. 2007;42.
**463.** Baird B, Charles A, Honeyman M, Maguire D, Das P. Understanding pressures in general practice: King's Fund London; 2016.
**464.** Lilly. World Alzheimer's Day Initial Report. Press release: Eli Lilly & Company; 2012.
**465.** Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia. 2018;14(4):535-562.
**466.** Bunnik EM, Richard E, Milne R, Schermer MH. On the personal utility of Alzheimer's disease-related biomarker testing in the research context. Journal of medical ethics. 2018:medethics-2018-104772.
**467.** Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ : British Medical Journal. 2006;332(7549):1089-1092.
**468.** Bayer AJ. The role of biomarkers and imaging in the clinical diagnosis of dementia. Age and Ageing. 2018;47(5):641-643.
**469.** Sheehan B. Assessment scales in dementia. Therapeutic advances in neurological disorders. 2012;5(6):349-358.
**470.** Croft P, Altman DG, Deeks JJ, et al. The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. BMC medicine. 2015;13(1):20.
**471.** Pickett J, Bird C, Ballard C, et al. A roadmap to advance dementia research in prevention, diagnosis, intervention, and care by 2025. International journal of geriatric psychiatry. 2018.
**472.** Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clinical chemistry. 2012;58(12):1636-1643.

**473.** Knottnerus JA, van Weel C, Muris JW. Evidence base of clinical diagnosis: evaluation of diagnostic procedures. BMJ: British Medical Journal. 2002;324(7335):477.
**474.** Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. Bmj. 2002;324(7338):669-671.
**475.** Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Clinical chemistry and laboratory medicine. 2003;41(1):68-73.
**476.** Walker R, Ratcliffe J, White A, Visvanathan R. Dementia assessment services: What are the perceptions of older people? Australasian journal on ageing. 2018;37(1):43-47.
**477.** Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials. 2017;18(1):122.
**478.** Gomes M, Pennington M, Wittenberg R, Knapp M, Black N, Smith S. Cost-effectiveness of Memory Assessment Services for the diagnosis and early support of patients with dementia in England. Journal of health services research & policy. 2017;22(4):226-235.
**479.** Michalowsky B, Flessa S, Hertel J, et al. Cost of diagnosing dementia in a German memory clinic. Alzheimer's research & therapy. 2017;9(1):65.