Neuropsychological tests for the diagnosis of Alzheimer's disease dementia and other dementias: a generic protocol for cross-sectional and delayed-verification studies

Daniel HJ Davis¹, Sam T Creavin², Anna Noel-Storr³, Terry J Quinn⁴, Nadjia Smailagic⁵, Chris Hyde⁶, Carol Brayne¹, Rupert McShane³, and Sarah Cullum⁷

¹Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

²School of Social and Community Medicine, University of Bristol, Bristol, UK

³Radcliffe Department of Medicine, University of Oxford, Oxford, UK

⁴Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

⁵Institute of Public Health, University of Cambridge, Cambridge, UK

⁶Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, University of Exeter, Exeter, UK

⁷Academic Unit of Psychiatry, School of Social and Community Medicine, Bristol University, Bristol, UK

Abstract

This is the protocol for a review and there is no abstract. The objectives are as follows:

- To determine the cross-sectional diagnostic accuracy of [index test] at various thresholds for ADD and other dementias [target condition] in [target population].

OR

- To determine the accuracy of [index test] at various thresholds for diagnosing ADD and other dementias [target condition] in [target population] after a follow-up period (delayed-verification studies).
- To investigate the heterogeneity of test accuracy in the included studies.
- To highlight the quality and quantity of research evidence available on the effectiveness of the index test in the target population.
- To identify gaps in the evidence and determine where further research is required.

Copyright © 2013 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Contact address: Daniel HJ Davis, Department of Public Health and Primary Care, University of Cambridge, Forvie Site, Robinson Way, Cambridge, CB2 0SR, UK. dhjd2@cam.ac.uk.

DECLARATIONS OF INTEREST

None.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.

Publication status and date: New, published in Issue 3, 2013.

BACKGROUND

Who is this document for?

This document has been created to aid authors writing Cochrane Dementia and Cognitive Improvement Group (CDCIG) Diagnostic Test Accuracy (DTA) protocols of neuropsychological tests. Each required heading of the protocol is described below with examples from ongoing CDCIG protocols provided to help explain the rationale for doing the review and the methods to be adopted. Authors may use the text from the examples provided and are encouraged to do so, adapting where necessary for variables such as index test and clinical setting. Links to appropriate sections of the *DTA Handbook* are also provided.

Context

The Cochrane Dementia and Cognitive Improvement Group (CDCIG, Oxford), in collaboration with the Cochrane Diagnostic Test Accuracy Review Unit (DTA Unit, Birmingham) and the Institute of Public Health (IPH, Cambridge), is working on a series of Cochrane Diagnostic Test Accuracy (DTA) reviews in dementia, which will be part of the National Institute for Health Research (NIHR) Cochrane Collaboration Programme. Appendix 1 shows the current content and structure of the programme, from which authors can see where their review fits.

Why is it important to evaluate the accuracy of diagnostic tests in dementia?

The revision of the clinical criteria for Alzheimer's disease dementia proposed by the National Institute on Aging and the US Alzheimer's Association widened the scope for biomarkers (such as brain imaging and cerebrospinal fluid analysis) to contribute to diagnostic categories (Jack 2011; McKhann 2011). However, the accuracy of biomarkers in the diagnosis of Alzheimer's disease dementia and other dementias has not yet been systematically evaluated. Clinical properties of dementia biomarkers should not be assumed; therefore, formal systematic evaluation of sensitivity, specificity, and other properties of biomarkers should be performed and collated in Cochrane DTA reviews. To ensure a comprehensive review of tests used in the assessment of possible dementia, the diagnostic accuracy of several of the neuropsychological tests and scales will be evaluated. Once these individual reviews have been completed, we plan to undertake a review of the comparative and incremental value of all included tests in the diagnosis of Alzheimer's disease dementia and, if evidence is sufficient, other dementias.

The aim of this generic protocol is to provide a framework for authors writing Cochrane DTA protocols for evaluation of the accuracy of neuropsychological tests in the diagnosis of dementias.

DTA review title structure

See DTA Handbook Chapter 4, Section 4.2.

Title—Title: [Index test] *for the diagnosis of* [Target condition] *in* [Target population].

See: DTA Handbook Chapter 4, Table 4.2a, option 3

Examples

- [Example 1] Mini Mental State Examination [Index test] for the diagnosis of Alzheimer's disease dementia and other dementias [Target condition] in people aged 65 in the general population [Target population].

In Example 1, the review evaluates the accuracy of the index test in the diagnosis of Alzheimer's disease dementia and other dementias in older people with or without memory problems who are living in the community (akin to population screening). The reference standard would be one of the clinical definitions, as applied by an expert(s). Here the time between applications of the index test and the reference standard is ideally as short as possible, although it is often a matter of weeks in dementia studies. This is the standard cross-sectional study design employed for the evaluation of diagnostic test accuracy.

- [Example 2] Addenbrookes Cognitive Examination [Index test] for the diagnosis of Alzheimer's disease dementia and other dementias [Target condition] in people with mild cognitive impairment (MCI) aged 65 and older [Target population].

In Example 2, the review evaluates the ability of the index test to distinguish which of the people already diagnosed with mild cognitive impairment (MCI) at study baseline will develop a full dementia syndrome. In this example, the included studies are likely to have baseline study populations (of participants with MCI) ascertained and followed up in a memory clinic, usually within a secondary care setting, although it would also be possible to identify and follow up a group of people with MCI in population-based cohort studies. This study design can be described as a 'delayed-verification study' because it includes a necessary period of followup to determine those who have a diagnosis of dementia. These studies are sometimes described as longitudinal or predictive, but note that terms such as 'longitudinal prediction' or 'longitudinal diagnosis' will not be used in the titles of CDCIG DTA reviews. See DTA Handbook Chapter 6, Section 6.1 (for more on delayed cross-sectional studies)

Target condition being diagnosed

Dementia is a progressive syndrome of global cognitive impairment. In the UK, it affects 5% of the population over 65 years of age and at least 25% of those over 85 years of age (Alzheimer's Society 2007). Worldwide, 36 million people were estimated to be living with dementia in 2010 (Wimo 2010), and this number will increase to more than 115 million by 2050, with the greatest increases in prevalence predicted to occur in the developing regions. China and its Western Pacific neighbours are predicted to have 26 million people living with dementia by 2040 (Ferri 2005). Dementia encompasses a group of disorders characterised by progressive loss of both cognitive function and ability to perform activities of daily living that can be accompanied by neuropsychiatric symptoms and challenging behaviours of varying type and severity (Appendix 2). The underlying pathology is usually degenerative, and subtypes of dementia include Alzheimer's disease dementia (ADD), vascular dementia,

dementia with Lewy bodies, and frontotemporal dementia. However, considerable overlap may be noted in the clinical and pathological presentations (CFAS 2001). For example, Alzheimer's pathology is commonly present in the brain of people with dementia with prominent vascular or Lewy body pathology. Similarly, those with prominent Alzheimer's pathology also commonly have coexisting vascular or Lewy body pathology (Matthews 2009; Savva 2009). We aim to assess the accuracy of commonly used neuropsychological tests in detecting ADD and other dementias.

The number of people with dementia is increasing, and it is becoming more important to make an accurate diagnosis early in the clinical process, which will enable optimal treatment and planning of care. However, no *in vivo* reference standard has been universally agreed upon for diagnosis of the dementias, and even the value of diagnoses based on neuropathological criteria has been questioned (Scheltens 2011). This makes a systematic review of diagnostic test accuracy studies particularly challenging. As the central characteristic of dementia is progression of the disorder, some of the DTA reviews included in this series have chosen to define the target condition as the development of ADD or another dementing illness in people with mild cognitive impairment (MCI), arguing that evidence of continued cognitive and functional decline is a more clinically valid marker of pathology that gives rise to dementia than is a cross-sectional "snap shot" diagnosis. The target condition will be dementia or its subtypes (as defined by the reference standards described later) identified either at the same time that the index test is administered (cross-sectional studies) or after some time has passed (delayed-verification studies).

Mild cognitive impairment (MCI)—MCI is a heterogeneous condition; over the past two decades, nearly 20 different classifications for an intermediate cognitive state have been proposed. In this protocol, the term 'MCI' refers to any of these definitions of MCI, including the clinical criteria put forth by Petersen et al (Petersen 1999; Petersen 2004; Winblad 2004) and (the closely related) newer criteria from the National Institute on Aging - Alzheimer's Association (NIA-AA) (McKhann 2011). Over time, people with MCI may experience a gradually progressive cognitive decline and changes in personality and behaviour. When the cognitive impairment in memory, reasoning, language, or visuospatial abilities interferes with daily function, patients are diagnosed with dementia. Research studies (Petersen 1999; Bruscoli 2004) indicate that an annual average of 10% to 15% of patients with MCI may progress to dementia, in particular ADD, but with wide variation depending upon the source of study participants. Self-selected clinic attendees have the highest conversion rates (Mitchell 2009).

Four outcomes have been noted among those within an MCI population: progression to ADD, progression to another dementia, maintenance of stable MCI, and recovery. At the present time, it is not possible to determine exactly which of those patients with MCI will convert to ADD or other dementias, and which will recover or remain stable. It is thus the aim of some of the DTA reviews for diagnostic test accuracy in dementia to evaluate the accuracy of the index tests in predicting those people with MCI who will progress to the full clinical syndrome of dementia.

Index test(s)

For future authors: Describe the index test briefly, plus normative data and threshold values in different populations, cite references for studies demonstrating validity and reliability of the index test for the target population in question.

Many individual neuropsychological tests are available. The aim of this series of DTA reviews is to evaluate those neuropsychological tests that are commonly used by clinicians as brief assessments of cognition in a range of settings to aid detection of dementia disorders.

The following tests will be evaluated:

- The Mini Mental State Examination (MMSE) (Folstein 1975) is a 30-question assessment of cognitive function that assesses attention and orientation, memory, registration, recall, calculation, language, and ability to draw a complex polygon. It takes around 7.3 minutes to administer this test to a person with dementia and around 5.6 minutes to administer this test to a person with normal cognition (Borson 2000). The MMSE has recently been subject to copyright restrictions (de Silva 2010).
- The Addenbrooke's Cognitive Examination - Revised (ACE-R) (Mioshi 2006) has been proposed as a screening tool that can be used to identify mild dementia; it may be capable of distinguishing between Alzheimer's disease and frontotemporal dementia, although this has not been proven (Crawford 2012). The ACE-R incorporates the MMSE and assesses attention, orientation, fluency, language, visuospatial function, and memory, yielding subscale scores for each domain. The maximum ACE-R score is 100, and it takes 20 minutes to administer the test (Brown 2009).
- The Mini-Cog is a two part screening test that is used to assess cognitive impairment (Borson 2000). Participants are told three items and are asked to remember them; they then are asked to draw a clock, and then to recall the three items. The test takes three minutes to administer.
- The Montreal Cognitive Assessment (MOCA) (Nasreddine 2005) is a 30-item test that takes 10 minutes to administer (Ismail 2010). It assesses short-term memory, visuospatial functioning, executive functioning, attention, concentration and working memory, language, and orientation. The MOCA is described as identifying people with mild cognitive impairment who are not identified by the MMSE (Nasreddine 2005). A version is available online.

The [index test] with scoring rules is available as Appendix 3 (a/b/c for different versions).

See DTA Handbook Chapter 4, Section 4.5

Alternative test(s)

CDCIG is in the process of conducting a series of DTA reviews of biomarkers and other tests to determine their sensitivity and specificity for the diagnosis of Alzheimer's disease

dementia and other dementias. This review includes tests that are available in any setting and tests that are available only in specialised care. Tests that may be available in a primary healthcare setting include:

- Informant interviews.
- Other tests of cognitive function.

Tests given for the diagnosis of Alzheimer's disease dementia and other dementias that are not available in primary care include:

- PET-FDG (positron emission tomography F-fluorodeoxyglucose)
- PET-PiB (positron emission tomography-Pittsburgh compound B)
- sMRI (structural magnetic resonance imaging)
- CSF (cerebrospinal fluid analysis of abeta and tau)
- APOE e4 (apolipoprotein E e4)
- FP-CIT (fluoropropyl-carbomethoxy-3b-(4-iodophenyl)tropane) SPECT (single photon emission computed tomography)

Clinical pathway

Presentation—*For future authors: Describe the role of the test for a given target population, for example using the MMSE to case-find for dementia in general hospital settings. It may be informative to readers of your protocol to include a diagram of care pathways. You may also wish to discuss possible uses of the index test in clinical practice or health policy initiatives (e.g. such as population screening), should its diagnostic accuracy be acceptable.*

Dementia develops over several years. It is presumed that people are asymptomatic for a period, during which pathology is accumulating. Individuals or their relatives may then notice subtle impairments in recent memory. Gradually, more cognitive domains become involved, and difficulty planning complex tasks becomes increasingly apparent. In the UK, people usually present to their general practitioner, who may administer the index tests and will potentially refer them to a hospital memory clinic. However, many people with dementia do not present until much later in the disorder and will follow a different pathway to diagnosis, for example, being identified during admission to a general hospital for a physical illness. Thus the pathway influences the accuracy of the diagnostic test. The accuracy of the test will vary with the experience of the administrator, and the accuracy of the subsequent diagnosis will vary with the history of referrals to the particular healthcare setting. Diagnostic assessment pathways may vary among countries, and diagnoses may be made by a variety of specialists, including neurologists, psychiatrists, and geriatricians.

Standard diagnostic practice—Standard assessment of dementia includes history taking, clinical examination (including neurological, mental state, and cognitive examinations), and an interview with a relative or other informant. Before dementia is diagnosed, the clinician should identify and if possible treat other physical and mental

disorders, such as hypothyroidism or depression, that might be contributing to cognitive impairment. A neuroradiological examination (computed tomography (CT) or magnetic resonance imaging (MRI) of the brain) is recommended in most recent guidelines (McKhann 2011;NICE 2006), but sometimes the diagnosis is made on the basis of history and presentation alone.

Dementia as diagnosed is defined by a deficit in more than two cognitive domains of sufficient degree to impair functional activities. Symptoms are usually progressive over a period of at least several months and should not be attributable to any other brain disorder. General diagnostic criteria for dementia as provided in the International Classification of Diseases, 10th edition (ICD-10), and the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), are detailed in Appendix 2. In some settings, a two-stage screening and assessment process takes place. Screening of people with suspected dementia usually requires a brief test of cognitive function and/or informant questionnaires; a low score reveals those people who require more in-depth assessment (Boustani 2003). The diagnostic accuracy of some of these screening tests will be evaluated in this series of Cochrane DTA reviews.

How might the index test improve diagnoses, treatments, and patient

outcomes?—Accurate diagnosis leads to opportunities for treatment. At the present time, no cure for dementia is known, but treatments are available that can slow cognitive and functional decline, or reduce the associated behavioural and psychiatric symptoms of dementia (Birks 2006; Clare 2003; McShane 2006). Furthermore, if ADD (and other dementias) can be diagnosed at an early stage, people with dementia, their families, and potential caregivers will be able to make timely plans for the future. Coupled with appropriate contingency planning, proper recognition of the disorder may also help to prevent inappropriate and potentially harmful admissions to hospital or institutional care (Bourne 2007). In addition, accurate and early identification of dementia may provide opportunities for the use of newly evolving interventions designed to delay or prevent progression to more debilitating stages of dementia. Improved diagnostic accuracy might also reduce false positive diagnoses, which carry a risk of significant costs (in the form of further unnecessary investigations or treatment) and harm (from side effects of investigations, treatment, or anxiety).

Prior tests—Most tests (e.g. neuroimaging, cerebrospinal fluid (CSF) analysis) are performed after a cognitive deficit is noted. However, it is conceivable that patients with abnormalities on brain imaging, which may be performed for any number of reasons, are more likely to be tested subsequently for cognitive deficits.

Rationale

The public health burden of cognitive and functional impairment due to dementia is a matter of growing concern. With the changing age structure of populations in both high- and low-income countries, overall dementia prevalence is increasing (Ferri 2005). At the population level, this increased prevalence has major implications for service provision and planning, given that the condition leads to progressive functional dependence over several years. In

the UK, it is estimated that the annual expenditure on dementia care is £23 billion (Luengo-Fernandez 2012), and the worldwide cost of dementia in 2010 was USD604 billion (Wimo 2010). Accurate and early diagnosis is crucial for planning care. In addition, accurate diagnosis is critical if participants for adequately powered clinical trials are to be identified. This generic protocol will consider the diagnostic accuracy of neuropsychological tests for Alzheimer's disease dementia, along with other dementias. Definitions of dementia will be based on clinical standards with or without neuropathological confirmation.

Scope of generic protocol—This generic protocol serves as a guide to specified methods for the systematic review of the diagnostic accuracy of several neuropsychological tests as used in defined target populations. We expect to identify both cross-sectional and delayed-verification study designs (see Table 1). Our aim is to produce separate reviews for cross-sectional and delayed-verification studies for each index test in each defined population. The characteristics of the target population will determine the healthcare settings in which the studies take place. For example, people who self-present with memory problems are likely to be “captured” in a primary care setting, whereas those who require a neuropsychological test battery to investigate cognitive impairment are more likely to have been referred to a secondary care setting such as a memory clinic. In practice, studies may be conducted across different population groups, may vary in the representation of MCI, and may even combine cross-sectional and delayed-verification methods. If the number and range of studies identified are limited to only a few across different study populations, it may be preferable to present the findings in a single review; in this case, a meta-analysis would not be carried out because inherent test performance is not the same across different populations, and findings would have variable implications for policy and practice.

OBJECTIVES

To determine the cross-sectional diagnostic accuracy of [index test] at various thresholds for ADD and other dementias [target condition] in [target population].

OR

To determine the accuracy of [index test] at various thresholds for diagnosing ADD and other dementias [target condition] in [target population] after a follow-up period (delayed-verification studies).

Secondary objectives

- To investigate the heterogeneity of test accuracy in the included studies.
- To highlight the quality and quantity of research evidence available on the effectiveness of the index test in the target population.
- To identify gaps in the evidence and determine where further research is required.

METHODS

Criteria for considering studies for this review

Types of studies—Neuropsychological tests can be used in the diagnosis of dementia in the context of two main study designs: cross-sectional and longitudinal. Investigators using cross-sectional study designs administer the index test to all study participants, who are simultaneously assessed for diagnosis by an expert or by a trained researcher using a standardised diagnostic interview. In longitudinal study designs, participants do not meet the criteria for diagnosis at baseline but are followed up for expert case ascertainment in the future (these studies are known as “delayed verification of diagnosis”; see *DTA Handbook*, Chapter 6). As many of the dementing illnesses are thought to start several years before clinical diagnosis, the latter design is considered to be appropriate for DTA reviews of dementia. These two approaches to diagnosis verification can be applied to three different types of study populations: complete cohort; cases and controls selected from the same study population cohort (known as a ‘nested case-control study’); and cases and controls selected from undefined, and possibly unrelated, populations (known as a ‘classic case-control study’; see Table 1). All of these are associated with differing degrees of potential spectrum bias.

If nested case-control studies or cohort studies are available, then classic case-control studies should be excluded because they are based on an inferior study design that overestimates sensitivity and specificity. However, in the absence of nested case-control studies and cohort studies, classic case-control studies may be presented in the review as the current best available evidence for diagnostic test accuracy of a specific index test, but the inherent bias of this design should be acknowledged. In case-control studies, the controls are people without the target condition, which introduces the problem of spectrum bias. Diagnostic test accuracy may be overestimated if the controls are very healthy (and performing at the top of the range on cognitive tests) compared with the cases. Thus, it should be emphasised that reviews of only case-control studies will produce biased estimates of sensitivity and specificity; for this reason, it may be best to avoid performing a meta-analysis to produce a summary estimate in the review, especially because such summary estimates might be used inappropriately by decision-makers. Alternatively, reviewers may decide to state that no reliable research can be found.

In nested case-control studies, both cases and controls are selected from a defined cohort and are usually subject to less spectrum bias than in classic case-control studies (depending on how recruitment was carried out). In a cohort study, all subjects in the cohort are examined, so this study design has the least spectrum bias and provides the best generalisable evidence for diagnostic accuracy, provided the study population can be shown to be truly representative of the target population. Most commonly used neuropsychological tests are neither expensive nor invasive; thus we are likely to find sufficient evidence from cohort or nested case-control studies, and it is unlikely that we would need to consider classic case-control study designs.

In reviews of delayed-verification studies, we will consider all longitudinal (or prospective) studies of participants without dementia at baseline, in which progression to a dementia is

included as an outcome in a sample that represents a defined population, including prospective cohort and nested case-control studies. We will also consider participants in the control groups of randomised controlled trials (these study designs can be defined as modified prospective studies in which an intervention is randomly allocated to participants who have been selected to represent a defined population). It should be highlighted that delayed-verification studies will not include any individuals receiving active treatment for dementia. No minimum follow-up time will be stipulated, but follow-up periods will be investigated as a source of heterogeneity in the data analysis.

The lower right-hand box in Table 1 represents a case-control study design with delayed verification. This would be possible only if the index test on the selected cases and controls had fortuitously been carried out at an earlier stage. This might be possible for MMSE, for example, which tends to get done on everyone at an early stage, but the study would not have been designed this way (i.e. prospectively), or it would fall into the nested case-control category.

Settings—Performance of both cross-sectional and delayed-verification DTA studies will be assessed in specific settings that relate to the target population at a particular point in the diagnostic pathway. We recognise that primary care is defined differently around the world. We identify and define the following types of studies as possibly examining the utility of MMSE for the diagnosis of dementia:

- Community or population-based studies examine the diagnostic utility of MMSE in an unselected group of people, regardless of health or residential status (i.e. they need not be complaining of subjective memory complaints). Findings of reviews in this setting might be applicable to population-based dementia screening initiatives. Other community-based studies in which entry into the study does not fulfil criteria for population-based sampling frames (e.g. volunteer studies) may also be included here, as long as the population can be described, and may be considered as a source of heterogeneity.
- Primary care-based studies examine the diagnostic utility of MMSE in a group of people who are thought to have a memory complaint, as expressed by themselves or by a caregiver, family member, or friend. We define primary care as non-specialist, office-based care. Patients who are opportunistically identified by a formal care provider (e.g. nurse, home care worker) who subsequently refers the patient to primary care for further assessment would also be included in primary care-based studies.
- Secondary care-based studies examine the diagnostic utility of MMSE in a group of people who are thought to have a memory complaint, and who have been referred from another healthcare provider. We define secondary care as specialist care provided by an individual with an interest or advanced training in the area (e.g. psychologist, neurologist, geriatrician, psychiatrist). We consider office-based specialists to be an example of those providing secondary care. Studies conducted in the context of a specialist setting, even if it is the first contact the patient has had with healthcare provision, will be included in this group because results from these

studies will not be applicable to clinicians in primary health care. We will include outpatient and inpatient secondary care settings but recognise that this will be a source of heterogeneity.

Participants—We will include all participants who meet criteria for inclusion in [study population] and are thus considered representative of [target population]. We will exclude studies of participants with a secondary cause for cognitive impairment, namely, current or history of alcohol/drug abuse, central nervous system (CNS) trauma (e.g. subdural haematoma), tumour, or infection.

For delayed-verification studies in people with MCI, we will include participants with MCI as determined by defined criteria (Matthews 2008; Petersen 1999; Petersen 2004; Winblad 2004) but recognise that the effect of different population definitions must be examined in the analyses. Participants with a family history of Alzheimer's dementia may be more readily diagnosed with dementia, perhaps leading to verification bias. On this basis, studies that exclusively investigate those with a known genetic predisposition will be examined separately. Furthermore, studies specifically investigating early-onset dementia and any studies that include more than 30% of participants younger than 65 years of age will be examined separately.

Information on participant attrition will be recorded, and its relation to neuropsychological testing at baseline will be quantitatively assessed if possible.

Index tests—These are examples of tests that are covered by this protocol. This generic protocol can be used beyond these specific examples.

- MMSE
- ACE-R
- MOCA
- Mini-Cog

Neuropsychological testing can be administered in a number of different ways: pen and paper, computer, self-administered. The length of time required to complete these tests is also of interest to practitioners and decision-makers reading the review.

Comparator tests: No comparator has been identified for this review because, at present, no standard practice neuropsychological test is available for dementia diagnosis. A portfolio of DTA reviews for dementia is planned (see Appendix 1); once completed, it will be possible to carry out comparative reviews and to evaluate the added value of additional tests (including biomarkers) in different settings.

Target conditions—The target condition is any stage of dementia of any aetiology, but we expect to find studies that focus on Alzheimer's disease dementia, vascular dementia, Lewy body dementia, and frontotemporal dementia. However, findings will be appraised separately if studies examining dementias of differing aetiologies or differing stages have been extracted and are included in the review.

Reference standards—No in vivo reference standard is available for diagnosis of the dementias. This makes a systematic review of diagnostic test accuracy studies particularly challenging. In this review, the target condition will be dementia or its subtypes (as defined by the reference standards described below) identified at the same time that the index test is administered.

Clinical diagnosis will include all-cause (unspecified) dementia, as defined by any recognised diagnostic criteria, for example, Diagnostic and Statistical Manual of Diseases, Fourth Edition (APA 1994), and International Classification of Diseases, 10th edition (WHO 1992) (see Appendix 2).

In studies in which a reference standard refers to different criteria for dementia (e.g. McKhann 1984: unlikely, possible, probable, definite), we will consider people as having the disease if they are classified as having either probable or definite Alzheimer's dementia.

Alzheimer's dementia: The National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (McKhann 1984) have put forth the best antemortem clinical consensus 'reference standard' for Alzheimer's disease dementia, defining three antemortem groups: probable, possible, and unlikely Alzheimer's dementia. Newer criteria for Alzheimer's disease dementia introduced in 2011 include the use of biomarkers (such as brain imaging and cerebrospinal fluid analysis) to contribute to diagnostic categories (McKhann 2011). However, we will present in a separate category any studies that use these criteria and will test the findings in a sensitivity analysis.

Lewy body dementia: The reference standard for Lewy body dementia is the revised or earlier McKeith criteria (McKeith 1996; McKeith 2005).

Frontotemporal dementia: The reference standard for frontotemporal dementia is the Lund criteria (Lund 1994).

Vascular dementia: The reference standard for vascular dementia is the National Institute of Neurological and Communicative Disorders and Stroke and the Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINCDS-AIREN) criteria (Roman 1993).

Neuropathological criteria: Overlap between normal ageing, Alzheimer's disease, and other types of dementia is generally noted for neuropathological diagnoses. Autopsy-defined criteria include the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) (Mirra 1991) and the Regan Institute criteria (Newell 1999). Studies using solely neuropathological outcome definitions must be regarded separately. However, when such studies report dementia diagnoses by clinical criteria (blinded to pathological status), these studies may be considered with the others.

We recognise that different iterations of reference standards over time may not be directly comparable (e.g. DSM-III-R versus DSM-IV, ICD-9 versus ICD-10), and that the validity of

diagnoses may vary with the degree or manner in which the criteria have been operationalised (e.g. individual clinician versus algorithm versus consensus determination). Data on the method and application of the reference standard will be collected and, if considered to be a source of bias, will be examined as a source of heterogeneity. Although it is unlikely that a specific reference standard might favour particular index tests, there is the more general issue of incorporation bias, in which the reference standard is applied with knowledge of the index test because neuropsychological deficits are integral to the definition of dementia. This is less problematic in cross-sectional studies because the index test and the reference standard may be administered completely independently.

For delayed-verification reviews, we will be considering application of the reference standard over the interval between baseline and first follow-up. For follow-up studies of MCI cohorts, the target condition and the reference standards for dementia diagnoses remain as above. The minimum time interval for progression from MCI to ADD or other dementia has been provisionally defined by CDCIG as 9 months; however, if studies with shorter time intervals are extracted, these will be included if they meet other inclusion criteria, and the effect of time to diagnosis will be examined in the investigation of heterogeneity.

Search methods for identification of studies

See DTA Handbook Chapter 7.

As soon as the title has been registered, authors will be put in touch with the CDCIG Trials Search Co-ordinator who is currently developing generic search strategies for the protocol.

Electronic searches—We will search MEDLINE (OvidSP), EMBASE (OvidSP), BIOSIS (Ovid), Science Citation Index (ISI Web of Knowledge), PsycINFO (Ovid), and LILACS (Bireme). See Appendix 5 for a proposed draft strategy to be run in MEDLINE (OvidSP). Similarly structured search strategies will be designed using search terms appropriate for each database. Controlled vocabulary such as MeSH terms and Emtree will be used where appropriate. In the searches developed, no attempt will be made to restrict studies on the basis of sampling frame or setting. This approach is intended to maximise sensitivity and allow inclusion on the basis of population-based sampling to be assessed at screening (below, 'Selection of studies'). Search filters (collections of terms aimed at reducing the number need to screen) will not be used as an overall limiter because those published have not proved sensitive enough (Whiting 2011a). No language restriction will be applied to the electronic searches; translation services will be used as necessary. We will also request a search of the Cochrane Register of Diagnostic Test Accuracy Studies (hosted and maintained by the Cochrane Renal Group) and the specialised register of the CDCIG, ALOIS, which includes both intervention and diagnostic test accuracy studies in dementia.

Initial searches will be performed by a single researcher who has extensive experience in systematic reviewing.

Searching other resources—We will check the reference lists of all relevant papers for additional studies.

We will also search:

- MEDION Database (Meta-analyses van Diagnostisch Onderzoek) www.york.ac.uk/inst/crd/crddatabases.html,
- DARE (Database of Abstracts of Reviews of Effects) www.york.ac.uk/inst/crd/crddatabases.html.
- HTA Database (www.cochrane.org/about-us/evidence-based-health-care/webliography/books/hta, via the Cochrane Library).
- ARIF Database (Aggressive Research Intelligence Facility) www.arif.bham.ac.uk.

Relevant studies will be used in PubMed to search for additional studies using the Related Articles feature. Key studies will be examined in citation databases such as Science Citation Index and Scopus to ascertain any additional relevant studies. Some grey literature will be identified through Science Citation Index, which now includes conference proceedings. We will aim to access theses and PhD abstracts from institutions known to be involved in prospective dementia studies. We will also attempt to contact researchers involved in studies with possibly relevant but unpublished data. We will not perform handsearching because little published evidence describes the benefits of handsearching for reports of DTA studies (Glanville 2012).

Data collection and analysis

Selection of studies—The eligibility criteria (inclusion and exclusion) should be detailed here, according to the issues identified in the preceding section for the specific protocol.

Initially, studies will be selected from title and abstract screening undertaken by the study authors or by teams of experienced assessors. After this, the full paper for each potentially eligible study identified by the search will be located. These papers will be independently evaluated by at least two authors for inclusion or exclusion, after the sampling frame for each study has been assessed. Disagreements will be resolved by discussion. If this does not prove conclusive, the default position will be to include the study. The study selection process will be detailed in a PRISMA flow diagram.

Data extraction and management—Data on study characteristics will be extracted to a study-specific pro forma and will include data for assessment of quality and for investigation of heterogeneity, as described in Appendix 6. The pro forma will have been piloted against ten primary diagnostic studies.

Data will be extracted by two review authors. The results will be dichotomised if necessary and cross-tabulated in two-by-two tables of index test results (positive or negative) against the target disorder (positive or negative). The results will be extracted directly onto RevMan tables.

Assessment of methodological quality—See DTA Handbook Chapter 9.

The methodological quality of each study will be assessed using the QUADAS-2 tool (Whiting 2011), as recommended by The Cochrane Collaboration. This tool is made up of

four domains: patient selection, index test, reference standard, and patient flow (Appendix 7). Each domain is assessed in terms of risk of bias; the first three domains are also considered in terms of applicability. Operational definitions and anchoring statements describing the use of QUADAS-2 are detailed in Appendix 8.

Statistical analysis and data synthesis—See DTA Handbook Chapter 10.

The target condition comprises two categories: (1) dementia (not otherwise specified) and (2) dementia subtypes (Alzheimer's, vascular, Lewy body, etc). Studies may detail one or both outcomes. For each index test in a given study setting, each of these target conditions will merit a separate meta-analysis.

For all included studies (both cross-sectional and delayed-verification), the data in the two-by-two tables (showing the binary test results cross-classified with the binary reference standard) will be used to calculate sensitivities and specificities, along with their 95% confidence intervals. We will present individual study results graphically by plotting estimates of sensitivities and specificities in a forest plot and in receiver operating characteristic (ROC) space. These findings will be considered in light of the previous systematic assessment (using QUADAS-2) of the methodological quality of individual studies. We will use RevMan software for these descriptive analyses and to produce summary ROC curves. If more than one threshold is reported in an individual study, we will present the graphical findings for all thresholds reported. However, we will avoid inclusion of study data in the calculation of a summary statistic on more than one occasion (in the same setting) by using only the threshold, which is considered to be 'standard practice' for the target population in question. Studies will not be pooled across settings. If no standard practice is agreed upon for the index test and the target population in question, the optimal threshold will be used (i.e. the threshold nearest to the upper left corner of the ROC curve) in calculating the summary ROC curve in RevMan and for any subsequent meta-analyses; we recognise that this may lead to an overestimation of diagnostic accuracy (Leeftang 2008). The basis for the selection of a threshold will vary with each review and will be specified. This will be an important component of the discussion section.

For delayed-verification studies, we will apply the same DTA framework for analysis of a single test, ignoring any censoring that might have occurred. We acknowledge that such a reduction in the data may represent a significant oversimplification. We will therefore adopt an 'intention to diagnose' (ITD) approach as well. If possible, we will present what the result would be if all dropouts had developed dementia and if all dropouts had not developed dementia. To do this, we may need to assume that the proportion of positive and negative test results is the same among unknown and known participants.

We will perform meta-analyses on pairs of sensitivity and specificity if it is appropriate to pool the data. Once the relevant studies have been identified, it will be clear whether most studies have reported results with consistent thresholds. If so, a bivariate random-effects approach based on pairs of sensitivity and specificity using bivariate random effects may be appropriate (Reitsma 2005). This approach enables us to calculate summary estimates of sensitivity and specificity, while correctly dealing with the different sources of variation: (1)

imprecision by which sensitivity and specificity have been measured within each study; (2) variation beyond chance in sensitivity and specificity between studies; and (3) any correlation that might exist between sensitivity and specificity. Categorised covariates can be incorporated in the bivariate model to examine the effects of potential sources of bias and variation across subgroups of studies, as outlined in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*, Chapter 10. Because of the bivariate nature of the model, effects on sensitivity and specificity can be modelled separately. The results of the bivariate model can be processed to calculate likelihood ratios. Calculating (negative) predictive values requires an estimate of the prevalence, in addition to values of sensitivity and specificity. If summary likelihood ratios can be derived, we will calculate predictive values based on population-based estimates of age-specific prevalence to estimate pretest probability.

If different thresholds are reported, we will use hierarchical summary ROC (HSROC) models (see DTA Handbook Chapter 10). Model fit will be assessed with the use of likelihood ratio tests. We will use Stata software, version 12.1 (StataCorp LP, College Station, Texas), to carry out additional analyses using bivariate or HSROC approaches.

Investigations of heterogeneity

Investigation of sources of heterogeneity: Framework for likely sources of heterogeneity (most important sources marked with *):

- Index test.
 - Thresholds.
 - Technical features (including different versions of the test).
 - Operator characteristics (e.g. training).*
- Target disorder.
 - Reference standard/s used: DSM definition, ICD definition, NINDS-ARDRA or other classification, including pathological definitions; and operationalisation of these classifications (e.g. individual clinician/algorithm/consensus group).
 - Spectrum of target disorder (may also depend on study design).*
- Target population.
 - Age, sex, education.*
 - Other characteristics (e.g. ApoE status, definition and duration of MCI at baseline (if applicable)).
 - Prevalence in different settings.
 - Treatment, previous or current interventions.
- Study quality.
 - Types of studies: see below.

- Prior clinical information will increase accuracy of the index test.
- Time from index test to reference standard (measured in days or weeks for cross-sectional studies).
- Duration of follow-up (measured in years for delayed-verification studies).
- Loss to follow-up.

It is likely that only a handful of studies are sufficiently robust to be included in the meta-analyses; this fact will allow only one or two sources of heterogeneity to be explored (because of insufficient data). The most important sources of heterogeneity in this series of reviews are likely to be, in order of importance:

- Age,
- Sex,
- Education.
- Severity (or stage) of the target disorder.
- Operator characteristics (for index test and reference standard).

It may be that heterogeneity due to disease severity is addressed principally in the QUADAS-2 assessments of spectrum bias. Although it is likely to be important, training required for test administration may not be well reported or easily operationalised. See DTA Handbook Chapter 4, DTA Handbook Chapter 10. Heterogeneity will be investigated in the first instance (informally) through visual examination of forest plots of sensitivities and specificities and through visual examination of the ROC plot of the raw data. Depending on the number of studies available, we will include as many covariates in the regression analyses as possible - up to 10 studies per covariate. We recognise that it is likely that power will be insufficient to allow formal investigation of all possible sources of heterogeneity. However, if we identify further likely sources of heterogeneity, we will investigate these by sub-group analyses and, if data allow, will include them as covariates in the regression model, with assistance from the DTA UK Support Unit.

Sensitivity analyses—We will perform sensitivity analysis to determine the effect of excluding studies that are deemed to be at high risk of bias according to the QUADAS-2 checklist. Additionally, we will perform sensitivity analyses to determine the effect of excluding studies that were flagged as possibly being less appropriate for inclusion (when disagreement between authors could not be resolved). Primary analysis will include all studies; sensitivity analysis will exclude studies of low quality (high likelihood of bias) to determine whether the results are influenced by inclusion of lower-quality studies.

Assessment of reporting bias—Quantitative methods for exploring reporting bias are not well established for studies of DTA. Specifically, funnel plots of the diagnostic odds ratio (DOR) versus the standard error of this estimate will not be considered.

Acknowledgments

The authors would like to thanks Anne Eisinga and Ayesha Khan for their contribution to developing the search strategy.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- Wellcome Trust, UK.
Daniel Davis is supported by a Wellcome Trust Research Training Fellowship
- This work is supported by NIHR Cochrane DTA programme grant 10 4001 05, UK.

Appendix 1. Overview of diagnostic test accuracy reviews in dementia

	Setting				
Index test BIOMARKER/IMAGING	Community	Primary	Secondary	All settings	No. of reviews
Abeta					
Delayed-verification				x	1
Cross-sectional				x	1
Tau/abeta ratio					
Delayed-verification				x	1
Cross-sectional					
PET-PiB					
Delayed-verification				x	1
Cross-sectional					
PET-FDG					
Delayed-verification				x	1
Cross-sectional					
Structural MRI					
Delayed-verification				x	1
Cross-sectional					
FP-CiT SPECT					
Delayed-verification					
Cross-sectional				x	1
SPECT scan					
Delayed-verification					
Cross-sectional				x	1
APOE e4					
Delayed-verification	x	x	x		3
Cross-sectional					

Index test SCALES	Community	Primary	Secondary	All settings	No. of reviews
IQCODE					
Delayed-verification				x	1
Cross-sectional	x	x	x		3
AD8					
Delayed-verification	x	x	x		3
Cross-sectional	x	x	x		3
MMSE					
Delayed-verification				x	1
Cross-sectional	x	x	x		3
Mini-Cog					
Delayed-verification					
Cross-sectional	x	x	x		3
MOCA					
Delayed-verification					
Cross-sectional	x	x	x		3
ACER					
Delayed-verification					
Cross-sectional					

Appendix 2. Classification of dementia

World Health Organisation International Classification of Diseases-10

G1. Evidence of each of the following:

- A decline in memory, which is most evident in the learning of new information, although in more severe cases, the recall of previously learned information also may be affected. The impairment applies to both verbal and nonverbal material. The decline should be objectively verified by obtaining a reliable history from an informant, supplemented, if possible, by neuropsychological tests or quantified cognitive assessments.
- A decline in other cognitive abilities characterised by deterioration in judgement and thinking, such as planning and organising, and in the general processing of information. Evidence for this should be obtained when possible by interviewing an informant, supplemented, if possible, by neuropsychological tests or quantified objective assessments. Deterioration from a previously higher level of performance should be established.

G2. Preserved awareness of the environment during a period long enough to enable the unequivocal demonstration of G1. When episodes of delirium are superimposed, the diagnosis of dementia should be deferred.

G3. A decline in emotional control or motivation, or a change in social behaviour, manifest as at least one of the following:

- Emotional liability.
- Irritability.
- Apathy.
- Coarsening of social behaviour.

G4. For a confident clinical diagnosis, G1 should have been present for at least six months; if the period since the manifest onset is shorter, the diagnosis can only be tentative.

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision

- A. The development of multiple cognitive deficits manifested by both
 1. Memory impairment (impaired ability to learn new information or to recall previously learned information)
 2. One (or more) of the following cognitive disturbances:
 - Aphasia (language disturbance).
 - Apraxia (impaired ability to carry out motor activities despite intact motor function).
 - Agnosia (failure to recognize or identify objects despite intact sensory function).
 - Disturbance in executive functioning (i.e. planning, organizing, sequencing, abstracting).
- B. Each of the cognitive deficits in Criteria A1 and A2
 - Causes significant impairment in social or occupational functioning.
 - Represents a significant decline from a previous level of functioning.
- C. The deficits do not occur exclusively during the course of a delirium.
- D. The disturbance is not better accounted for by another Axis I disorder (e.g. major depressive disorder, schizophrenia).

Appendix 3. [index test] and scoring rules

This appendix should specify the test and scoring rules relevant to the specific DTA review

Appendix 4. Commonly used cognitive assessments / screening tools

Abbreviated Mental Test (AMT)

Mini-Mental State Examination (MMSE)

Modified Mini-Mental State Examination (3-MS)

Addenbrooke's Cognitive Examination-Revised (ACE-R)

Clock drawing test (CDT)

Montreal Cognitive Assessment (MoCA)

Blessed Information-Memory-Concentration Test (Blessed I-M-C)

Mini-Cog

GP-Cog

Test Your Memory (TYM)

St Louis University Mental Status examination (SLUMS)

6-item cognitive impairment test (6-CIT)

Appendix 5. Search strategy (Medline Ovid SP) run for specialised register (ALOIS)

The MEDLINE search uses the following concepts:

- A.** Specific neuropsychological tests
- B.** General terms (both free text and MeSH) for tests/testing/screening
- C.** Outcome: dementia diagnosis (unfocused MeSH with diagnostic sub-headings)
- D.** Condition of interest: dementia (general dementia terms both free text and MeSH - exploded and unfocused)
- E.** Methodological filter: NOT used to limit all search

Concept combination:

- 1. (A OR B) AND C
 - 2. (A OR B) AND D AND E
 - 3. A AND E
- = 1 OR 2 OR 3

Setting not included as a concept in the MEDLINE search as these terms generally are not indexed well or consistently. This means that the search has been kept deliberately sensitive by not restricting it to a particular setting.

The search strategy

- 1. "word recall".ti,ab.
- 2. ("7-minute screen" OR "seven-minute screen").ti,ab.
- 3. ("6 item cognitive impairment test" OR "six-item cognitive impairment test").ti,ab.
- 4. "6 CIT".ti,ab.
- 5. "AB cognitive screen".ti,ab.
- 6. "abbreviated mental test".ti,ab.

7. "ADAS-cog".ti,ab.
8. AD8.ti,ab.
9. "inform* interview".ti,ab.
10. "animal fluency test".ti,ab.
11. "brief alzheimer* screen".ti,ab.
12. "brief cognitive scale".ti,ab.
13. "clinical dementia rating scale".ti,ab.
14. "clinical dementia test".ti,ab.
15. "community screening interview for dementia".ti,ab.
16. "cognitive abilities screening instrument".ti,ab.
17. "cognitive assessment screening test".ti,ab.
18. "cognitive capacity screening examination".ti,ab.
19. "clock drawing test".ti,ab.
20. "deterioration cognitive observee".ti,ab.
21. ("Dem Tect" OR DemTect).ti,ab.
22. "object memory evaluation".ti,ab.
23. "IQCODE".ti,ab.
24. "mattis dementia rating scale".ti,ab.
25. "memory impairment screen".ti,ab.
26. "minnesota cognitive acuity screen".ti,ab.
27. "mini-cog".ti,ab.
28. "mini-mental state exam*".ti,ab.
29. "mmse".ti,ab.
30. "modified mini-mental state exam".ti,ab.
31. "3MS".ti,ab.
32. "neurobehavio?ral cognitive status exam*".ti,ab.
33. "cognistat".ti,ab.
34. "quick cognitive screening test".ti,ab.
35. "QCST".ti,ab.
36. "rapid dementia screening test".ti,ab.
37. "RDST".ti,ab.
38. "repeatable battery for the assessment of neuropsychological status".ti,ab.

39. "RBANS".ti,ab.
40. "rowland universal dementia assessment scale".ti,ab.
41. "rudas".ti,ab.
42. "self-administered gerocognitive exam*".ti,ab.
43. ("self-administered" and "SAGE").ti,ab.
44. "self-administered computerized screening test for dementia".ti,ab.
45. "short and sweet screening instrument".ti,ab.
46. "sassi".ti,ab.
47. "short cognitive performance test".ti,ab.
48. "syndrome kurztest".ti,ab.
49. ("six item screener" OR "6-item screener").ti,ab.
50. "short memory questionnaire".ti,ab.
51. ("short memory questionnaire" and "SMQ").ti,ab.
52. "short orientation memory concentration test".ti,ab.
53. "s-omc".ti,ab.
54. "short blessed test".ti,ab.
55. "short portable mental status questionnaire".ti,ab.
56. "spmsq".ti,ab.
57. "short test of mental status".ti,ab.
58. "telephone interview of cognitive status modified".ti,ab.
59. "tics-m".ti,ab.
60. "trail making test".ti,ab.
61. "verbal fluency categories".ti,ab.
62. "WORLD test".ti,ab.
63. "general practitioner assessment of cognition".ti,ab.
64. "GPCOG".ti,ab.
65. "Hopkins verbal learning test".ti,ab.
66. "HVLt".ti,ab.
67. "time and change test".ti,ab.
68. "modified world test".ti,ab.
69. "symptoms of dementia screener".ti,ab.
70. "dementia questionnaire".ti,ab.

71. "7MS".ti,ab.
72. ("concord informant dementia scale" or CIDS).ti,ab.
73. (SAPH or "dementia screening and perceived harm*").ti,ab.
74. or/1-73
75. exp Dementia/
76. Delirium, Dementia, Amnestic, Cognitive Disorders/
77. dement*.ti,ab.
78. alzheimer*.ti,ab.
79. AD.ti,ab.
80. ("lewy bod*" or DLB or LBD or FTD or FTLD or "frontotemporal lobar degeneration" or "frontaltemporal dement*").ti,ab.
81. "cognit* impair*".ti,ab.
82. (cognit* adj4 (disorder* or declin* or fail* or function* or degenerat* or deteriorat*)).ti,ab.
83. (memory adj3 (complain* or declin* or function* or disorder*)).ti,ab.
84. or/75-83
85. exp "sensitivity and specificity"/
86. "reproducibility of results"/
87. (predict* adj3 (dement* or AD or alzheimer*)).ti,ab.
88. (identif* adj3 (dement* or AD or alzheimer*)).ti,ab.
89. (discriminat* adj3 (dement* or AD or alzheimer*)).ti,ab.
90. (distinguish* adj3 (dement* or AD or alzheimer*)).ti,ab.
91. (differenti* adj3 (dement* or AD or alzheimer*)).ti,ab.
92. diagnos*.ti.
93. di.fs.
94. sensitivit*.ab.
95. specificit*.ab.
96. (ROC or "receiver operat*").ab.
97. Area under curve/
98. ("Area under curve" or AUC).ab.
99. (detect* adj3 (dement* or AD or alzheimer*)).ti,ab.
- 100.sROC.ab.

- 101.accura*.ti,ab.
- 102.(likelihood adj3 (ratio* or function*)).ab.
- 103.(conver* adj3 (dement* or AD or alzheimer*)).ti,ab.
- 104.((true or false) adj3 (positive* or negative*)).ab.
- 105.((positive* or negative* or false or true) adj3 rate*).ti,ab.
- 106.or/85-105
- 107.exp dementia/di
- 108.Cognition Disorders/di [Diagnosis]
- 109.Memory Disorders/di
- 110.or/107-109
- 111.*Neuropsychological Tests/
- 112.*Questionnaires/
- 113.Geriatric Assessment/mt
- 114.*Geriatric Assessment/
- 115.Neuropsychological Tests/mt, st
- 116.“neuropsychological test*”.ti,ab.
- 117.(neuropsychological adj (assess* or evaluat* or test*)).ti,ab.
- 118.(neuropsychological adj (assess* or evaluat* or test* or exam* or battery)).ti,ab.
- 119.Self report/
- 120.self-assessment/ or diagnostic self evaluation/
- 121.Mass Screening/
- 122.early diagnosis/
- 123.or/111-122 124.
- 124.74 or 123
- 125.110 and 124
- 126.74 or 123
- 127.84 and 106 and 126
- 128.74 and 106
- 129.125 or 127 or 128
- 130.exp Animals/not Humans.sh.
- 131.129 not 130

The searches will identify a large number of citations to screen. However, we will employ the use of a team of trained screeners to work through the large numbers.

Appendix 6. Information for extraction to pro forma

A. Bibliographic details of primary paper:

- Author, title of study, year, and journal.

B. Details of index test:

- Method of [index test] administration, including who administered and interpreted the test, and their training.
- Thresholds used to define positive and negative tests.

C. Reference standard:

- Reference standard used.
- Method of [reference standard] administration, including who administered the test and their training.

D. Study population:

- Number of subjects.
- Age.
- Gender.
- Other characteristics (e.g. ApoE status).
- Settings: community, primary care, secondary care outpatients, and secondary care inpatients and residential care.
- Participant recruitment.
- Sampling procedures.
- Time between index test and reference standard.
- Proportion of people in sample with dementia.
- Subtype and stage of dementia if available.
- MCI definition used (if applicable).
- Duration of follow-up in delayed verification studies.
- Attrition and missing data.

Appendix 7. Assessment of methodological quality QUADAS-2

DOMAIN	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING
Description	Describe methods of patient selection: Describe included patients (prior testing, presentation, intended use of index test, and setting):	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index test(s) and/or reference standard, or who were excluded from the 2×2 table (refer to flow diagram): Describe the time interval and any interventions between index test (s) and reference standard:
Signalling questions (yes/no/unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive the same reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive a reference standard? Were all patients included in the analysis?
Risk of bias (high/ low/ unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability (high/ low/ unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation may differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Appendix 8. Anchoring statements for quality assessment of [index test] diagnostic studies

We provide some core anchoring statements for quality assessment of diagnostic test accuracy reviews of neuropsychological tests in dementia (see Appendix 9). These statements are designed for use with the QUADAS-2 tool and were derived during a two-day, multidisciplinary focus group in 2010. If a QUADAS-2 signalling question for a specific domain is answered 'yes,' the risk of bias can be judged to be 'low.' If a question is answered 'no,' this indicates a risk of potential bias. The focus group was tasked with judging the extent of the bias for each domain. During this process, it became clear that

certain issues were key to assessing quality, whilst others were important to record but were less important for assessing overall quality. To assist, we describe a 'weighting' system. When an item is weighted 'high risk,' that section of the QUADAS-2 results table is judged to have a high potential for bias if a signalling question is answered 'no.' For example, in dementia diagnostic test accuracy studies, ensuring that clinicians performing dementia assessment are blinded to results of the index test is fundamental. If this blinding was not present, the item on the reference standard should be scored 'high risk of bias,' regardless of the other contributory elements. When an item is weighted 'low risk,' it is judged to have a low potential for bias if a signalling question for that section of the QUADAS-2 results table is answered 'no.' Overall bias will be judged on whether other signalling questions (with a high risk of bias) for the same domain are also answered 'no.'

In assessing individual items, the score of 'unclear' should be given only if there is genuine uncertainty. In these situations, review authors will contact the relevant study teams for additional information.

Anchoring statements to assist with assessment for risk of bias

Domain 1: Patient selection

Risk of bias: could the selection of patients have introduced bias? (high/low/unclear)

Was a consecutive or random sample of patients enrolled?: When sampling is used, the methods least likely to cause bias are consecutive sampling and random sampling, which should be stated and/or described. Nonrandom sampling or sampling based on volunteers is more likely to be at high risk of bias.

Weighting: High risk of bias

Was a case-control design avoided?: Case-control study designs have a high risk of bias, but sometimes they are the only studies available, especially if the index test is expensive and/or invasive. Nested case-control designs (systematically selected from a defined population cohort) are less prone to bias, but they will still narrow the spectrum of patients that receive the index test. Study designs (both cohort and case-control) that may also increase bias are those designs in which the study team deliberately increases or decreases the proportion of subjects with the target condition, for example, a population study may be enriched with extra dementia subjects from a secondary care setting.

Weighting: High risk of bias

Did the study avoid inappropriate exclusions?: The study will be automatically graded as unclear if exclusions are not detailed (pending contact with study authors). When exclusions are detailed, the study will be graded as 'low risk' if review authors feel that the exclusions are appropriate. Certain exclusions common to many studies of dementia are medical instability, terminal disease, alcohol/substance misuse, concomitant psychiatric diagnosis, and other neurodegenerative condition. However, if 'difficult to diagnose' groups are excluded, this may introduce bias, so exclusion criteria must be justified. For a community

sample, we would expect relatively few exclusions. Post hoc exclusions will be labelled 'high risk' of bias.

Weighting: High risk of bias

Applicability: are there concerns that the included patients do not match the review question? (high/low/unclear)—The included patients should match the intended population as described in the review question. If not already specified in the review inclusion criteria, the setting will be particularly important - the review authors should consider population in terms of symptoms, pretesting, and potential disease prevalence. Studies that use very selected subjects or subgroups will be classified as low applicability, unless they are intended to represent a defined target population, for example, people with memory problems referred to a specialist and investigated by lumbar puncture.

Domain 2: Index test

Risk of bias: could the conduct or interpretation of the index test have introduced bias? (high/low/unclear)

Were the index test results interpreted without knowledge of the reference standard?:

Terms such as 'blinded' or 'independently and without knowledge of' are sufficient, and full details of the blinding procedure are not required. This item may be scored as 'low risk' if it is explicitly described, or if there is a clear temporal pattern to the order of testing that precludes the need for formal blinding (e.g. all [neuropsychological test] assessments were performed before the dementia assessment). As most neuropsychological tests are administered by a third party, knowledge of dementia diagnosis may influence their ratings; tests that are self-administered, for example, by using a computerised version, may have less risk of bias.

Weighting: High risk

Were the index test thresholds pre-specified?: For neuropsychological scales, there is usually a threshold above which subjects are classified as 'test positive'; this may be referred to as threshold, clinical cut-off, or dichotomisation point. Different thresholds are used in different populations. A study is classified at higher risk of bias if the authors define the optimal cut-off post hoc based on their own study data. Certain papers may use an alternative methodology for analysis that does not use thresholds, and these papers should be classified as not applicable.

Weighting: Low risk

Were sufficient data on (neuropsychological test) application given for the test to be repeated in an independent study?: Particular points of interest include method of administration (e.g. self-completed questionnaire versus direct questioning interview), nature of informant, and language of assessment. If a novel form of the index test is used, for example, a translated questionnaire, details of the scale should be included and a reference given to an appropriate descriptive text, and evidence of validation should be provided.

Weighting: Low risk

Applicability: are there concerns that the index test, its conduct, or its interpretation may differ from the review question? (high/low/unclear)—

Variations in the length, structure, language, and/or administration of the index test may all affect applicability if they differ from those specified in the review question.

Domain 3: Reference standard

Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias? (high/low/unclear)

Is the reference standard likely to correctly classify the target condition?: Commonly used international criteria that can assist with clinical diagnosis of dementia include those detailed in DSM-IV and ICD-10. Criteria specific to dementia subtypes include but are not limited to NINCDS-ADRDA criteria for Alzheimer's dementia; McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; and NINDS-AIREN criteria for vascular dementia. When the criteria used for assessment are not familiar to the review authors and the Cochrane Dementia and Cognitive Improvement Group, this item should be classified as 'high risk of bias.'

Weighting: High risk

Were the reference standard results interpreted without knowledge of the results of the index test?: Terms such as 'blinded' or 'independent' are sufficient, and full details of the blinding procedure are not required. This may be scored as 'low risk' if explicitly described, or if a clear temporal pattern to the order of testing is evident (e.g. all dementia assessments performed before (neuropsychological test) testing).

Informant rating scales and direct cognitive tests present certain problems. It is accepted that informant interview and cognitive testing are usual components of clinical assessment for dementia; however, specific use of the scale under review in the clinical dementia assessment should be scored as high risk of bias.

Weighting: High risk

Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study?: Particular points of interest for dementia assessment include the training/expertise of the assessor, whether additional information (e.g. neuroimaging; other neuropsychological test results) was available to inform the diagnosis, and whether this was available for all participants.

Weighting: Variable risk, but high risk if method of dementia assessment not described

Applicability: are there concerns that the target condition as defined by the reference standard does not match the review question? (high/low/unclear)—

There is the possibility that some methods of dementia assessment, although valid, may

diagnose a smaller or larger proportion of subjects with disease than in usual clinical practice. In these instances, the item should be rated 'poor applicability.'

Domain 4: Patient flow and timing (N.B. refer to, or construct, a flow diagram)

Risk of bias: could the patient flow have introduced bias? (high/low/unclear)

Was there an appropriate interval between the index test and the reference standard?:

For a cross-sectional study design, there is potential for the subject to change between assessments; however, dementia is a slowly progressive disease that is not reversible. The ideal scenario would be a same-day assessment, but longer periods of time (e.g. several weeks or months) are unlikely to lead to a high risk of bias. For delayed-verification studies, the index and reference tests are necessarily separated in time, given the nature of the condition.

Weighting: Low risk

Did all subjects receive the same reference standard?: In some scenarios, subjects who score 'test positive' on the index test have a more detailed assessment for the target condition. When dementia assessment (or the reference standard) differs between subjects, this should be classified as high risk of bias.

Weighting: High risk

Were all subjects included in the final analysis?: Attrition will vary with study design. Delayed-verification studies will have higher attrition than cross-sectional studies because of mortality, and this is likely to be greater in subjects with the target condition. Dropouts (and missing data) should be accounted for. Attrition that is higher than expected (compared with other similar studies) should be treated as a high risk of bias. We have defined a cut-off of greater than 20% attrition as being high risk, but this will be highly dependent on the length of follow-up in individual studies.

Weighting: High risk

REFERENCES

Additional references

* *Indicates the major publication for the study*

Alzheimer's Society 2007. Alzheimer's Society. Dementia UK: A report into the prevalence and cost of dementia prepared by the Personal Social Services Research Unit (PSSRU) at the London School of Economics and the Institute of Psychiatry at King's College London, for the Alzheimer's Society. Alzheimer's Society; 2007.

APA 1994. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-IV). American Psychiatric Association; Washington DC: 1994.

Birks 2006. Birks J. Cholinesterase inhibitors for Alzheimer's disease. Cochrane Database of Systematic Reviews. 2006; 3 Art. No.: CD005593.

Borson 2000. 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; 15:1021-7. [PubMed: 11113982]

- Bourne 2007. Bourne, J. Improving services and support for people with dementia. National Audit Office; UK: 2007.
- Boustani 2003. 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. *Annals of Internal Medicine*. 2003; 138:927–37. [PubMed: 12779304]
- Brown 2009. Brown J, Pengas G, Dawson K, Brown LA, Clatworthy P. Self administered cognitive screening test (TYM) for detection of Alzheimer's disease: cross sectional study. *BMJ*. 2009; 338:b2030. [PubMed: 19509424]
- Bruscoli 2004. Bruscoli M, Lovestone S. Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics*. 2004; 16:129–40. [PubMed: 15318760]
- CFAS 2001. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. *Lancet*. 2001; 357:169–75. [PubMed: 11213093]
- Clare 2003. Clare L, Woods RT, Moniz Cook ED, Orrell M, Spector A. Cognitive rehabilitation and cognitive training for early-stage Alzheimer's disease and vascular dementia. *Cochrane Database of Systematic Reviews*. 2003; 4:CD003260. [PubMed: 14583963]
- Crawford 2012. Crawford S, Whitnall L, Robertson J, Evans JJ. 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. *International Journal of Geriatric Psychiatry*. 2012; 27:659–69. [PubMed: 22068971]
- de Silva 2010. de Silva V, Hanwella R. Why are we copyrighting science? *BMJ*. 2010; 341:c4738. [PubMed: 20847026]
- DTA Handbook Chapter 10. Macaskill, P.; Gatsonis, C.; Deeks, JJ.; Harbord, R.; Yemisi Takwoingi, Y.; The Cochrane Collaboration. Chapter 10: Analysing and presenting results. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 2010. <http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Chapter%2010%20%20Version%201.0.pdf><http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Chapter%2010%20%20Version%201.0.pdf>
- DTA Handbook Chapter 4. Smidt, N.; Deeks, J.; Moore, T.; The Cochrane Collaboration. Chapter 4: Guide to the contents of a Cochrane review and protocol. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 2010. http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Handbook1CC_Chapter04_101.pdfhttp://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Handbook1CC_Chapter04_101.pdf
- DTA Handbook Chapter 6. Bossuyt, PM.; Leeftang, MM.; The Cochrane Collaboration. Chapter 6: Developing criteria for including studies. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 2008. <http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Chapter06-Including-Studies%20%28September-2008%29.pdf><http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Chapter06-Including-Studies%20%28September-2008%29.pdf>
- DTA Handbook Chapter 7. de Vet, HCW.; Eisinga, A.; Riphagen, II.; Aertgeerts, B.; Pewsner, D.; The Cochrane Collaboration. Chapter 7: Searching for Studies. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 2008. <http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Chapter07-Searching-%28September-2008%29.pdf><http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/Chapter07-Searching-%28September-2008%29.pdf>
- DTA Handbook Chapter 9. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeftang MMG, Deeks JJ, The Cochrane Collaboration. Chapter 9: Assessing methodological quality. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 2009http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/ch09_Oct09.pdf
- Ferri 2005. Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, et al. Global prevalence of dementia: a Delphi consensus study. *Lancet*. 2005; 366:2112–7. [PubMed: 16360788]
- Folstein 1975. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State - Practical method for grading cognitive state of patients for clinicians. *Journal of Psychiatric Research*. 1975; 12:189–98. [PubMed: 1202204]

- Glanville 2012. Glanville JM, Cikalo M, Crawford F, Dozier M, McIntosh H. Handsearching did not yield additional unique FDG-PET diagnostic test accuracy studies compared with electronic searches: a preliminary investigation. *Research Synthesis Methods*. Sep; 2012 3(3):202–13.
- Ismail 2010. Ismail Z, Rajji TK, Shulman KI. Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry*. 2010; 25:111–20. [PubMed: 19582756]
- Jack 2011. Jack CR Jr, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*. 2011; 7:257–62.
- Leefflang 2008. 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; 54:729–37. [PubMed: 18258670]
- Luengo-Fernandez 2012. Luengo-Fernandez R, Leal J, Gray AM. UK research expenditure on dementia, heart disease, stroke and cancer: are levels of spending related to disease burden? *European Journal of Neurology*. 2012; 19:149–54. [PubMed: 21951976]
- Lund 1994. The Lund and Manchester Groups. Clinical and neuropathological criteria for frontotemporal dementia. *Journal of Neurology, Neurosurgery, and Psychiatry*. 1994; 57:416–8.
- Matthews 2008. Matthews FE, Stephan BC, McKeith IG, Bond J, Brayne C. Two-year progression from mild cognitive impairment to dementia: to what extent do different definitions agree? *Journal of the American Geriatrics Society*. 2008; 56:1424–33. [PubMed: 18662209]
- Matthews 2009. Matthews FE, Brayne C, Lowe J, McKeith I, Wharton SB, Ince P. Epidemiological pathology of dementia: attributable-risks at death in the Medical Research Council Cognitive Function and Ageing Study. *PLoS Medicine*. 2009; 6:e1000180. [PubMed: 19901977]
- McKeith 1996. McKeith IG, Galasko D, Kosaka K, Perry EK, Dickson DW, Hansen LA, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the Consortium on DLB International workshop. *Neurology*. 1996; 47:1113–24. [PubMed: 8909416]
- McKeith 2005. McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology*. 2005; 65:1863–72. [PubMed: 16237129]
- McKhann 1984. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984; 34:939–44. [PubMed: 6610841]
- McKhann 2011. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*. 2011; 7:263–9.
- McShane 2006. McShane R, Areosa Sastre A, Minakaran N. Memantine for dementia. *Cochrane Database of Systematic Reviews*. 2006; 3:CD003154.
- Mioshi 2006. 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; 21:1078–85. [PubMed: 16977673]
- Mirra 1991. Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*. 1991; 41:479–86. [PubMed: 2011243]
- Mitchell 2009. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica*. 2009; 119:252–65. [PubMed: 19236314]
- Nasreddine 2005. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*. 2005; 53:695–9. [PubMed: 15817019]

- Newell 1999. Newell KL, Hyman BT, Growdon JH, Hedley-Whyte ET. Application of the National Institute on Aging (NIA)-Reagan Institute criteria for the neuropathological diagnosis of Alzheimer disease. *Journal of Neuropathology and Experimental Neurology*. 1999; 58:1147–55. [PubMed: 10560657]
- NICE 2006. Dementia: supporting people with dementia and their careers in health and social care. National Institute for Health and Clinical Excellence; 2006. Vol. Clinical Guideline 42
- Petersen 1999. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology*. 1999; 56:303–8. [PubMed: 10190820]
- Petersen 2004. Petersen RC. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*. 2004; 256:183–94. [PubMed: 15324362]
- Reitsma 2005. 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. *Journal of Clinical Epidemiology*. 2005; 58:982–90. [PubMed: 16168343]
- Roman 1993. Roman GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*. 1993; 43:250–60. [PubMed: 8094895]
- Savva 2009. Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C. Age, neuropathology, and dementia. *New England Journal of Medicine*. 2009; 360:2302–9. [PubMed: 19474427]
- Scheltens 2011. Scheltens P, Rockwood K. How golden is the gold standard of neuropathology in dementia? *Alzheimer's & Dementia*. 2011; 7:486–9.
- Whiting 2011. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*. 2011; 155:529–36. [PubMed: 22007046]
- Whiting 2011a. Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *Journal of Clinical Epidemiology*. 2011a; 64:602–7. [PubMed: 21075596]
- WHO 1992. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. WHO; Geneva, Switzerland: 1992. Geneva: World Health Organization
- Wimo 2010. Wimo, A.; Prince, M. World Alzheimer's Report 2010: the global economic impact of dementia. Alzheimer's Disease International (ADI); London, UK: 2010. London, UK
- Winblad 2004. Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, et al. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*. 2004; 256:240–6. [PubMed: 15324367]

Table 1
Study designs to assess diagnostic test accuracy in dementia categorised by origin of participants and the timing of verification

		Origin of population		
		Cohort (a defined population containing a typical number and spectrum of the target condition)	Cases and controls chosen from a defined cohort	Cases and controls chosen from undefined (and possibly unrelated) populations
Timing of verification	Index test and reference standard simultaneous	'Cohort with cross-sectional verification study'	'Nested case-control with cross-sectional verification study' (<i>index test done at time of diagnosis</i>)	'Classic case-control with cross-sectional verification study'
	Reference standard applied after an extended period	'Cohort with delayed-verification study'	'Nested case-control with delayed-verification study' (<i>index test planned for baseline, before diagnosis</i>)	'Classic case-control with delayed-verification study' (<i>index test done earlier than diagnosis, fortuitous rather than planned</i>)

Increasing strength of evidence for DTA (right to left).