Cochrane Library

Cochrane Database of Systematic Reviews

Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings (Review)

Harrison JK, Stott DJ, McShane R, Noel-Storr AH, Swann-Price RS, Quinn TJ

Harrison JK, Stott DJ, McShane R, Noel-Storr AH, Swann-Price RS, 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 2016, Issue 11. Art. No.: CD011333.
DOI: 10.1002/14651858.CD011333.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings

Jennifer K Harrison[1], David J Stott[2], Rupert McShane[3], Anna H Noel-Storr[3], Rhiannon S Swann-Price[4], Terry J Quinn[2]

[1]Centre for Cognitive Ageing and Cognitive Epidemiology and the Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [2]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [3]Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [4]Dental School, University of Glasgow, Glasgow, UK

Contact address: Terry J Quinn, Institute of Cardiovascular and Medical Sciences, University of Glasgow, New Lister Building, Glasgow Royal Infirmary, Glasgow, Strathclyde, G4 0SFR, UK. Terry.Quinn@glasgow.ac.uk.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.
Publication status and date: New, published in Issue 11, 2016.

Citation: Harrison JK, Stott DJ, McShane R, Noel-Storr AH, Swann-Price RS, 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* 2016, Issue 11. Art. No.: CD011333. DOI: 10.1002/14651858.CD011333.pub2.

ABSTRACT

Background

The Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE) is a structured interview based on informant responses that is used to assess for possible dementia. IQCODE has been used for retrospective or contemporaneous assessment of cognitive decline. There is considerable interest in tests that may identify those at future risk of developing dementia. Assessing a population free of dementia for the prospective development of dementia is an approach often used in studies of dementia biomarkers. In theory, questionnaire-based assessments, such as IQCODE, could be used in a similar way, assessing for dementia that is diagnosed on a later (delayed) assessment.

Objectives

To determine the diagnostic accuracy of IQCODE in a population free from dementia for the delayed diagnosis of dementia (test accuracy with delayed verification study design).

Search methods

We searched these sources on 16 January 2016: ALOIS (Cochrane Dementia and Cognitive Improvement Group), MEDLINE Ovid SP, Embase Ovid SP, PsycINFO Ovid SP, BIOSIS Previews on Thomson Reuters Web of Science, Web of Science Core Collection (includes Conference Proceedings Citation Index) on Thomson Reuters Web of Science, CINAHL EBSCOhost, and LILACS BIREME. We also searched sources specific to diagnostic test accuracy: MEDION (Universities of Maastricht and Leuven); DARE (Database of Abstracts of Reviews of Effects, in the Cochrane Library); HTA Database (Health Technology Assessment Database, in the Cochrane Library), and ARIF (Birmingham University). We checked reference lists of included studies and reviews, used searches of included studies in PubMed to track related articles, and contacted research groups conducting work on IQCODE for dementia diagnosis to try to find additional studies. We developed a sensitive search strategy; search terms were designed to cover key concepts using several different approaches run in parallel, and included terms relating to cognitive tests, cognitive screening, and dementia. We used standardised database subject headings, such as MeSH terms (in MEDLINE) and other standardised headings (controlled vocabulary) in other databases, as appropriate.

Selection criteria

We selected studies that included a population free from dementia at baseline, who were assessed with the IQCODE and subsequently assessed for the development of dementia over time. The implication was that at the time of testing, the individual had a cognitive problem sufficient to result in an abnormal IQCODE score (defined by the study authors), but not yet meeting dementia diagnostic criteria.

Data collection and analysis

We screened all titles generated by the electronic database searches, and reviewed abstracts of all potentially relevant studies. Two assessors independently checked the full papers for eligibility and extracted data. We determined quality assessment (risk of bias and applicability) using the QUADAS-2 tool, and reported quality using the STARDdem tool.

Main results

From 85 papers describing IQCODE, we included three papers, representing data from 626 individuals. Of this total, 22% (N = 135/626) were excluded because of prevalent dementia. There was substantial attrition; 47% (N = 295) of the study population received reference standard assessment at first follow-up (three to six months) and 28% (N = 174) received reference standard assessment at final follow-up (one to three years). Prevalence of dementia ranged from 12% to 26% at first follow-up and 16% to 35% at final follow-up.

The three studies were considered to be too heterogenous to combine, so we did not perform meta-analyses to describe summary estimates of interest. Included patients were poststroke (two papers) and hip fracture (one paper). The IQCODE was used at three thresholds of positivity (higher than 3.0, higher than 3.12 and higher than 3.3) to predict those at risk of a future diagnosis of dementia. Using a cut-off of 3.0, IQCODE had a sensitivity of 0.75 (95%CI 0.51 to 0.91) and a specificity of 0.46 (95%CI 0.34 to 0.59) at one year following stroke. Using a cut-off of 3.12, the IQCODE had a sensitivity of 0.80 (95%CI 0.44 to 0.97) and specificity of 0.53 (95C%CI 0.41 to 0.65) for the clinical diagnosis of dementia at six months after hip fracture. Using a cut-off of 3.3, the IQCODE had a sensitivity of 0.84 (95%CI 0.68 to 0.94) and a specificity of 0.87 (95%CI 0.76 to 0.94) for the clinical diagnosis of dementia at one year after stroke.

In generaI, the IQCODE was sensitive for identification of those who would develop dementia, but lacked specificity. Methods for both excluding prevalent dementia at baseline and assessing for the development of dementia were varied, and had the potential to introduce bias.

Authors' conclusions

Included studies were heterogenous, recruited from specialist settings, and had potential biases. The studies identified did not allow us to make specific recommendations on the use of the IQCODE for the future diagnosis of dementia in clinical practice. The included studies highlighted the challenges of delayed verification dementia research, with issues around prevalent dementia assessment, loss to follow-up over time, and test non-completion potentially limiting the studies. Future research should recognise these issues and have explicit protocols for dealing with them.

PLAIN LANGUAGE SUMMARY

Using a structured questionnaire (the IQCODE) to detect individuals who may go on to develop dementia

Background

Accurately identifying people with dementia is an area of public and professional concern. Dementia is often not diagnosed until late in the disease, and this may limit timely access to appropriate health and social support. There is a growing interest in tests that detect dementia at an early stage, before symptoms have become problematic or noticeable. One way to do this is to test a person and then re-assess them over time to see if they have developed dementia.

Our review focused on the accuracy of a questionnaire-based assessment for dementia, called the IQCODE (Informant Questionnaire for Cognitive Decline in the Elderly). We described whether the initial IQCODE score can identify people who will develop dementia months or years after their first IQCODE assessment.

We searched electronic databases of published research studies, looking for all studies that looked at IQCODE and a later diagnosis of dementia. We searched from the first available papers in scientific databases up to and including January 2016.

Study characteristics

We found three relevant studies, all of which were carried out in specific hospital settings. Two papers only included patients with acute stroke, and the other included those who had sustained a hip fracture. The papers differed in many other ways, so we we were unable to estimate a summary of their combined results. In general, a 'positive' IQCODE picked up patients who would go on to develop dementia (good sensitivity), but mislabelled a number who did not develop dementia (poor specificity). We cannot make recommendations for current practice, based on the studies we reviewed.

Quality of the evidence

The included studies demonstrated some of the challenges of research that follows people at risk of dementia over time. Not all the studies had a robust method of ensuring that none of the included participants had dementia at the start of the study, and that only new cases were identified. Similarly, many of the participants included at the start of the study were not available for re-assessment, due to death or other illness.

The review was performed by a team based in research centres in the UK (Glasgow, Edinburgh, Oxford). We had no external funding specific to this study, and we have no conflicts of interest that may have influenced our assessment of the research data.

BACKGROUND

Dementia is a substantial and growing public health concern (Herbert 2013; Prince 2013). Depending on the case definition used, contemporary estimates of dementia prevalence in the United States are in the range of 2.5 to 4.5 million individuals. Changes in population demographics will be accompanied by increases in global dementia incidence and prevalence. Although the magnitude of the increase in prevalent dementia is debated, there is no doubt that absolute numbers of older adults with dementia will increase substantially in the short to medium-term future (Ferri 2005).

A diagnosis of dementia requires both cognitive and functional decline. A syndrome of cognitive problems beyond those expected for age and education, but not sufficient to impact on daily activities is also recognised. This possible intermediate state between normal cognitive ageing and pathological change is often labelled as mild cognitive impairment (MCI) or cognitive impairment, no dementia (CIND), although a variety of other terms are also used. For consistency, we use the term MCI throughout this review. A proportion of individuals with MCI will develop a clinical dementia state over time (estimated at 10% to 15% of MCI individuals annually), while others will improve or remain stable. All definitions of this 'pre-dementia' state are based on key criteria of changes in cognition (subjective or reported by an informant) with objective cognitive impairment, but preserved functional ability.

A key element of effective management in dementia is early, robust diagnosis. Recent guidelines place emphasis on very early diagnosis to facilitate improved management, and to allow informed discussions and planning with patients and carers (Cordell 2013). An early or unprompted assessment paradigm needs to distinguish early pathological change from normal states. Diagnosis of early dementia or MCI is especially challenging. It is important to recognise those who will progress to dementia, as identification of this group may allow for targeted intervention. However, at present, there is no accepted method for determining prognosis.

The ideal would be expert, multidisciplinary assessment, informed by various supplementary investigations (neuropsychology, neuroimaging or other biomarkers). This approach is only really feasible in a specialist memory service and is not suited to population screening or case-finding.

In practice, a two-stage process is often used, with initial triage assessments that are suitable for use by non-specialists used to select those patients who require further detailed assessment (Boustani 2003). Various tools for initial cognitive screening have been described (Brodaty 2002; Folstein 1975; Galvin 2005). Regardless of the methods used, there is room for improvement, as observational work suggests that many patients with dementia are not diagnosed (Chodosh 2004; Valcour 2000).

The initial assessment often takes the form of brief, direct cognitive testing. Such an approach will only provide a snapshot of cognitive function. However, a defining feature of dementia is cognitive or neuropsychological change over time. Patients themselves may struggle to make an objective assessment of personal change, and so an attractive approach is to question collateral sources with sufficient knowledge of the patient. These informant-based interviews aim to retrospectively assess change in function. An instrument that is prevalent in research and clinical practice, particularly in Europe, is the Informant Questionnaire for Cognitive Decline in

the Elderly (IQCODE) with questionnaire-based interviews. This screening or triage tool is the focus of this review (Jorm 2004).

Traditional assessment tools for cognitive problems have defined threshold scores that differentiate individuals likely to have dementia from those with no dementia. As dementia is a progressive, neurodegenerative disease, a population with cognitive problems will have a range of test scores. Individuals with a pre-dementia state, MCI, or indeed early dementia, may have screening test scores that although not at a threshold suggestive of dementia, are still abnormal for age. It seems plausible that a subthreshold score on a screening test such as IQCODE could be predictive of future dementia states, and so could be used to target those individuals who may need follow-up or further investigation. This paradigm of using an outcome of delayed verification of a dementia state is commonly used in studies of the diagnostic properties of dementia biomarkers, but theoretically, can be applied to direct or informant-based assessment scales.

This review focused on the use of the IQCODE in individuals without a firm clinical diagnosis of dementia, and assessed the accuracy of IQCODE scores for delayed verification of a diagnosis of dementia after a period of prospective follow-up.

Target condition being diagnosed

The target condition for this diagnostic test accuracy review was the development of all cause dementia (incident clinical diagnosis). Dementia is a syndrome characterised by cognitive or neuropsychological decline, sufficient to interfere with usual functioning. The neurodegeneration and clinical manifestations of dementia are progressive.

Dementia remains a clinical diagnosis, based on history from the patient and suitable collateral sources, and direct examination, including cognitive assessment. There is no universally accepted, ante-mortem, gold standard diagnostic strategy. We have chosen expert clinical diagnosis as our gold standard (reference standard), as we believe this is most in keeping with current diagnostic criteria and best practice.

A diagnosis of dementia can be made according to various internationally accepted diagnostic criteria, with exemplars being the World Health Organization International Classification of Diseases (ICD) and the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (DSM) for all cause dementia and subtypes. The label of dementia encompasses varying pathologies, of which Alzheimer's disease is the most common. Diagnostic criteria are available for specific dementia subtypes, that is, the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for Alzheimer's dementia (McKhann 1984; McKhann 2011); the McKeith criteria for Lewy Body dementia (McKeith 2005); the Lund criteria for frontotemporal dementias (McKhann 2001); and the NINDS-AIREN criteria for vascular dementia (Roman 1993).

We examined delayed verification of dementia, and so we have described the properties of a standard, initial assessment (the IQCODE) for detection of problems earlier in the disease journey than frank dementia. Thus, our outcome of interest for this review is a confirmed diagnosis at a point in time later than the initial IQCODE testing. We did not pre-specify a minimum or maximum length of follow-up.

A proportion of participants included in relevant studies were likely to have MCI, that is, cognitive problems beyond those expected for age and education but not sufficient to impact on daily activities. The usual research definition of MCI is that described by Petersen (Peterson 2004); and various subtypes have been proposed within the rubric of MCI. We collated information on MCI described using any validated criteria, however, the focus of the review was not IQCODE for the contemporaneous diagnosis of MCI, but rather IQCODE for a future diagnosis of dementia. These two constructs are related but not synonymous, as only a proportion of individuals with MCI will develop dementia.

Index test(s)

Our index test was the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE (Jorm 1988)).

The IQCODE was originally described as a 26-item informant questionnaire that sought to retrospectively ascertain change in cognitive and functional performance over a 10-year time period. IQCODE was designed as a brief screen for potential dementia, usually administered as a questionnaire given to the relevant proxy. For each item, the chosen proxy scores change on a five-point ordinal hierarchical scale, with responses ranging from 1: 'has become much better' to 5: 'has become much worse'. This gives a sum score of 26 to 130 that can be averaged by the total number of completed items, to give a final score of 1.0 to 5.0, where higher scores indicate greater decline.

First described in 1989, use of the IQCODE is prevalent in both clinical practice and research. A literature describing the properties of IQCODE is available, including studies of non-English IQCODE translations, studies in specific patient populations, and modifications to the original 26-item direct informant interview (Isella 2002; Jorm 1989; Jorm 2004). Versions of the IQCODE have been produced in other languages including: Chinese, Dutch, Finnish, French, Canadian French, German, Italian, Japanese, Korean, Norwegian, Polish, Spanish, and Thai (www.anu.edu.au/iqcode/). A shortened 16-item version is also available; this modified IQCODE is common in clinical practice and has been recommended as the preferred IQCODE format (Jorm 2004). Further modifications to the IQCODE are described, including fewer items and assessment over shorter time periods. Our analysis included all versions of IQCODE, but results for original and modified scales were not pooled. In this review, the term IQCODE

refers to the original 26-item English language questionnaire as described by Jorm. Other versions of IQCODE are described according to the number of items and administration language (that is, a 16-item IQCODE for Spanish speakers is described as IQCODE-16 Spanish).

In the original IQCODE development and validation work, normative data were described, with a total score higher than 93 or an average score higher than 3.31 indicative of cognitive impairment (Jorm 2004). There is no consensus on the optimal threshold and certainly no guidance on the use of subthreshold IQCODE scores for delayed verification. In setting thresholds for any diagnostic test, there is a trade-off between sensitivity and specificity, with the preferred values partly determined by the purpose of the test.

This review completes a suite of Cochrane reviews describing the test accuracy of IQCODE in various health care contexts (Harrison 2014; Harrison 2015; Quinn 2014).

Clinical pathway

Dementia develops over a trajectory of several years and screening tests may be performed at different stages in the dementia pathway. In this review, we considered any use of IQCODE as an initial assessment for cognitive decline, and we did not limit studies to a particular healthcare setting. We operationalised the various settings where the IQCODE may be used as secondary care, primary care, and community.

In secondary care settings, individuals would have been referred for expert input, but not exclusively due to memory complaints. Opportunistic screening of adults presenting as unscheduled admissions to hospitals would be an exemplar secondary care pathway. The rubric of secondary care also included individuals referred to dementia and memory specific services. This population would have had a high prevalence of cognitive disorders and mimics. More individuals would have had a greater degree of prior cognitive assessment than in other settings, but cognitive testing was not always performed prior to memory service referral (Menon 2011).

In the general practice and primary care setting, the individual self-presented to a non-specialist service because of subjective memory complaints. Previous cognitive testing was unlikely, but prevalence would be reasonably high. Using IQCODE in this setting could be described as triage or case-finding. In the community setting, the cohort was largely unselected and the approach may be described as population screening.

The IQCODE delayed verification approach recognises that in any of these settings or pathways, there will be a population who do not yet have a cognitive syndrome that would warrant a dementia label, but who nonetheless may progress to a frank dementia state. If IQCODE has delayed verification utility, this population may score less than expected on initial IQCODE assessment.

The IQCODE is not a diagnostic tool and was not designed to be used as such. Rather, IQCODE would often be used as part of an initial assessment, and based on test scores, more detailed assessment may be required. However, in order to quantify the test accuracy of the IQCODE, it was necessary to evaluate it as a diagnostic test, against a gold standard of clinical diagnosis.

IQCODE is often used, and may have particular utility, as an initial assessment in a group of individuals considered to be at risk of having or developing dementia. Here, the role of IQCODE is identifying those who may need further detailed assessment or follow-up. Although this description does not fulfil all the established criteria to be considered a screening test (Wilson 1968), we used the term 'screening' in this review as a descriptor of this early triage assessment.

Alternative test(s)

Several other dementia screening and assessment tools have been described, for example, Folstein's mini-mental state examination (MMSE; Folstein 1975). These performance-based measures for cognitive screening all rely on comparing single or multi-domain cognitive testing against population-specific normative data.

Other informant interviews are also available. For example, the AD-8 is an eight-question tool that requires dichotomous responses (yes or no) and tests for perceived changes in memory, problem solving, orientation, and daily activities (Galvin 2005).

For this review, we focused on papers that described IQCODE diagnostic properties; we did not consider other cognitive screening or assessment tools. Our IQCODE diagnostic test accuracy studies form part of a larger body of work by the Cochrane Dementia and Cognitive Improvement Group that describes test properties of all commonly used assessment tools (Appendix 1).

Rationale

There is no consensus on the optimal initial assessment for dementia, and choice is currently dictated by experience with a particular instrument, time constraints, and training. A better understanding of the diagnostic properties of various strategies would allow for an informed approach to testing. Critical evaluation of the evidence base for screening tests or other diagnostic markers is of major importance. Without a robust synthesis of the available information, there is the risk that future research, clinical practice, and policy will be built on erroneous assumptions about diagnostic validity.

This review forms part of a body of work that describes the diagnostic properties of commonly used dementia tools. At present, we are conducting single test reviews and meta-analyses. However, the intention is then to collate these data by performing an overview, that will allow comparison of various test strategies.

OBJECTIVES

To determine the diagnostic accuracy of the informant-based questionnaire IQCODE in a population free from dementia, for the delayed diagnosis of dementia.

Secondary objectives

Where data were available, we planned to describe the following:
1. The delayed verification diagnostic accuracy of IQCODE at various thresholds. We recognise that various thresholds or cut-off' scores have been used to define IQCODE screen-positive states, and thus various subthreshold cut-points could be used to describe individuals with cognitive problems not diagnostic of dementia. We did not pre-specify IQCODE cut-points of interest, rather we collected delayed verification test accuracy data for all cut-points described in the primary papers.
2. Effects of heterogeneity on the reported diagnostic accuracy of IQCODE for delayed verification dementia (see below).
Items of specific interest included case-mix of population, IQCODE test format, time since index test, and healthcare setting.

METHODS

Criteria for considering studies for this review

Types of studies

In this review we looked at the properties of IQCODE for diagnosis of the dementia state on prospective follow-up, that is, investigating whether a certain score on IQCODE, that may or may not be below the normal threshold, in a population free of dementia at baseline assessment, is associated with the development of dementia over a period of follow-up. The implication was that at the time of testing, the individual had a cognitive problem sufficient to be picked up on screening, but not yet meeting diagnostic criteria for dementia. We described this paradigm as 'delayed verification' diagnostic test accuracy. Other Cochrane reviews covered IQCODE for contemporaneous diagnosis of dementia (Harrison 2014; Harrison 2015; Quinn 2014).
We anticipated that the majority of studies would be performed in secondary care settings. We included test studies performed in other healthcare settings, and classified these as primary care or community.
We did not include case-control studies, since they are known to potentially overestimate properties of a test.
We did not include case studies or samples with very small numbers (for the purposes of this review, fewer than 10 participants), but described them in the table of excluded studies.
There may be cases where settings were mixed, for example, a population study 'enriched' with additional cases from primary care. If available, we considered separate data for patients from each setting. If these data were not available, we treated these studies as case-control studies, and did not include them in this review.

Participants

All adults (aged over 18 years) and with no formal diagnosis of dementia were eligible.
We did not predefine exclusion criteria relating to the case-mix of the population studied, but assessed this aspect of the study as part of our assessment of heterogeneity. Where there was concern that the participants were not representative, we explored this at study level, using the 'Risk of bias' assessment framework, outlined below.

Index tests

Studies had to include (not necessarily exclusively) IQCODE as an informant questionnaire for delayed verification.
IQCODE has been translated into a number of languages to facilitate international administration (Isella 2002). The properties of a translated IQCODE in a cohort of non-English speakers may differ from properties of the original English language questionnaire. We collected data on the principal language used for IQCODE assessment.
For this review, we did not consider other cognitive screening or assessment tools. Where a paper described the IQCODE with an in-study comparison against another screening tool, we included the IQCODE data only. Where IQCODE was used in combination with another cognitive screening tool, we included the IQCODE data only.

Target conditions

We included any clinical diagnosis of all cause (unspecified) dementia. Defining a particular dementia subtype was not required, although, where available, these data were recorded.

Reference standards

Our reference standard was a clinical diagnosis of incident dementia. We recognise that clinical diagnosis itself has a degree of variability, but this is not unique to dementia studies and does not invalidate the basic diagnostic test accuracy approach. We also recognise the lack of an agreed 'gold standard' reference for dementia, but believe a clinical reference is most relevant to the review topic, and in keeping with current best practice in dementia accuracy research.
For our primary analysis, clinical diagnosis, we included all cause (unspecified) dementia, using any recognised diagnostic criteria (for example ICD-10, DSM-IV). A diagnosis of dementia may

specify a pathological subtype; we included all common dementia subtypes (for example, NINCDS-ADRDA, Lund-Manchester, McKeith, NINCDS-AIREN). We did not define preferred diagnostic criteria for rarer forms of dementia (for example, alcohol-related, HIV-related, prion disease-related), and we considered them under our rubric of 'all cause' dementia, rather than separately.
Clinicians may use imaging, pathology, or other data to aid diagnosis, however, we did not include diagnoses based only on these data, without a corresponding clinical assessment. We recognise that different iterations of diagnostic criteria may not be directly comparable, and that diagnoses may vary with the degree or manner in which the criteria have been operationalised (for example, individual clinician versus algorithm versus consensus determination); we collected data on the method and application of the diagnosis of dementia for each study, and explored potential effects as part of our assessment of risk of bias and generalisability. Use of other (brief) direct performance tests in isolation were not an acceptable method for diagnosis.
We recognise that the diagnosis of dementia often comprises a degree of informant assessment. Thus there was potential for incorporation bias. We explored the potential effects of this bias through our 'Risk of bias' assessment.

Search methods for identification of studies

We used a variety of information sources to ensure that we included all relevant studies. We devised terms for electronic database searching in conjunction with the Information Specialist at the Cochrane Dementia and Cognitive Improvement Group. As part of a body of work looking at cognitive assessment tools, we created a sensitive search strategy designed to capture papers about dementia test accuracy. We then assessed the output of the searches to select those papers that could be pertinent to IQCODE, with further selection for directly relevant papers, and those papers with a delayed-verification methodology.

Electronic searches

We searched ALOIS, the specialised register of the Cochrane Dementia and Cognitive Improvement Group (which includes both intervention and diagnostic accuracy studies), MEDLINE OvidSP, Embase OvidSP, PsycINFO OvidSP, BIOSIS Previews on Thomson Reuters Web of Science, Web of Science Core Collection (includes Conference Proceedings Citation Index) on Thomson Reuters Web of Science, CINAHL EBSCOhost, and LILACS BIREME. See Appendix 2 and Appendix 3 for the strategies run. The original search date was 28 January 2013, with an updated search performed on 16 January 2016.
We also searched sources specific to diagnostic accuracy and healthcare research assessment on 16 January 2016:

- MEDION database (Meta-analyses van Diagnostisch Onderzoek: www.mediondatabase.nl);
- DARE (Database of Abstracts of Reviews of Effects in the Cochrane Library);
- HTA Database (Health Technology Assessment Database in the Cochrane Library);
- ARIF database (Aggressive Research Intelligence Facility: www.arif.bham.ac.uk).

We applied no language or date restrictions to the electronic searches and used translation services as necessary.
A single researcher (ANS), with extensive experience of systematic reviews from the Cochrane Dementia and Cognitive Improvement Group, performed the initial screening of the search results. All subsequent searches of titles, abstracts, and papers were performed independently by paired assessors (TJQ, JKH & RSP).

Searching other resources

Grey literature: We identified grey literature by searching conference proceedings, theses, or PhD abstracts in Embasee, the Web of Science Core Collection, and other databases already specified.
Handsearching: We did not perform handsearching. The evidence for the benefits of handsearching are not well defined, and we noted that a study specific to diagnostic accuracy studies suggested little additional benefit of handsearching above a robust initial search strategy (Glanville 2012) .
Reference lists: We checked the reference lists of all included studies and reviews in the field for further possible titles, and repeated the process until we found no new titles (Greenhalgh 1997).
Correspondence: We contacted research groups who have published or are conducting work on IQCODE for the diagnosis of dementia, informed by results of the initial search.
We searched for studies in PubMed, using the 'related article' feature. We examined key studies in citation databases of Science Citation Index and Scopus to identify any further studies that could potentially be included.

Data collection and analysis

Selection of studies

The original search was done for the programme of reviews in 2013. One review author (ANS) screened all titles generated by the initial electronic database searches for relevance. The initial search was a sensitive, generic search, designed to include all potential dementia screening tools. Two review authors (ANS, TJQ) selected titles potentially relevant to IQCODE. Two authors in the IQCODE review group (TJQ, PF) independently conducted further review and selection from the long list. We reviewed potential IQCODE-related titles, assessing all eligible studies as abstracts, and assessed potentially relevant studies as full manuscripts against our inclusion criteria. We resolved disagreement by discussion, with

the potential to involve a third review author (DJS) as arbiter, if necessary. We adopted a hierarchical approach to exclusion, first excluding on the basis of index test and reference standard, and then on the basis of sample size and study data. A focused update search was performed in 2016, which sought to identify only IQCODE studies with a delayed verification design. Two review authors (TJQ, JKH) independently reviewed potential IQCODE-related titles from this update, assessed the abstracts of all potentially relevant studies, and the full manuscripts of eligible studies against the inclusion criteria. We resolved disagreement by discussion, with the potential to involve a third review author (DJS) as arbiter if necessary.

Both in the original search and the update, where a study may have included useable data but these were not presented in the published manuscript, or the data presented could not be extracted to a standard two-by-two table, we contacted the authors directly to request further information or source data. If authors did not respond, or if the data were not available, we did not include the study (labelled as 'data not suitable for analysis' on the study flowchart). If the same data set was presented in more than one paper, we included the primary paper. We detailed the study selection process in a PRISMA flow diagram.

Data extraction and management

We extracted data to a study-specific pro forma that included clinical and demographic details of the participants, details of the setting, details of IQCODE administration, and details of the dementia diagnosis process.

Test accuracy data were extracted to a standard two-by-two table. Two review authors (TJQ, JKH) independently extracted data. The review authors were based in different centres and were blinded to each other's data until extraction was complete. We then compared and discussed data pro formas with reference to the original papers, resolving disagreements in data extraction by discussion, with the potential to involve a third review author (DJS) as arbiter if necessary.

For each included paper, we detailed the flow of participants (numbers recruited, included, assessed) in a flow diagram.

Assessment of methodological quality

As well as describing test accuracy, an important goal of the diagnostic test accuracy (DTA) process is to improve study design and reporting in dementia diagnostic studies. For this reason, we assessed both methodological and reporting quality, using two complementary processes.

We assessed the quality of study reporting using the dementia-specific extension to the Standards for the Reporting of Diagnostic Accuracy studies (STARD-dem) checklist (Noel-Storr 2014; Appendix 4).

We assessed the methodological quality of each study, using the Quality Assessment tool for Diagnostic Accuracy Studies (QUADAS-2) tool (www.bris.ac.uk/quadas/quadas-2). This tool incorporates domains specific to patient selection, index test, reference standard, and participant flow. Each domain is assessed for risk of bias, and the first three domains are also assessed for applicability. Operational definitions describing the use of QUADAS-2 are detailed in Appendix 5. To create QUADAS-2 anchoring statements specific to studies of dementia test accuracy, we convened a multidisciplinary review of various test accuracy studies with a dementia reference standard (Davis 2013; Appendix 6).

Paired, independent raters (TJQ and JKH), blinded to each other's scores, performed both assessments. We resolved disagreements by further review and discussion, with the potential to involve a third review author (DJS) as arbiter if necessary.

We did not use QUADAS-2 data to form a summary quality score, but rather, we chose to present a narrative summary that described studies that found high, low, or unclear risk of bias or concerns regarding applicability, with corresponding tabular and graphical displays.

Statistical analysis and data synthesis

We were principally interested in the test accuracy of IQCODE for the delayed diagnosis of dementia using a dichotomous variable, 'dementia' or 'no dementia'. Thus, we applied the current DTA framework for analysis of a single test and fitted the extracted data to a standard two-by-two data table showing binary test results cross-classified with a binary reference standard. We repeated this process for each IQCODE threshold score described. We further repeated the process for each assessment where the reference standard was assessed at more than one follow-up.

Where data allowed, we used Review Manager 5.3 (RevMan 2014) to calculate sensitivity, specificity, and their 95% confidence intervals (CIs) from the two-by-two tables abstracted from the included studies, or using data supplied from authors. The delayed verification nature of the included studies added a further level of complexity as a proportion of individuals recruited at baseline may be lost to subsequent review, and the delayed verification assessment may be performed at varying times from the initial IQCODE assessment. In the first instance, we applied the usual DTA framework, describing common reference time points and performing no imputation or adjustment for any drop-outs that might have occurred. We acknowledge that such a reduction in the data may represent a significant oversimplification.

We presented data graphically, using forest plots to allow basic visual inspection and comparison of individual studies. Standard forest plots with graphical representation of summary estimates are not suited to quantitative synthesis of DTA data. If data allowed, we had planned to calculate summary estimates of test accuracy. In our protocol, we pre-specified that we would consider meta-analyses if more than three studies with suitable data were available. We planned to use the bivariate approach to give summary estimates of test accuracy at common thresholds and common

time points, and to use the HSROC model to explore differing thresholds across studies.

Investigations of heterogeneity

Heterogeneity is to be expected in DTA reviews, and we did not perform formal analysis to quantify heterogeneity.

We included IQCODE studies that spanned various settings and offered a narrative review of all studies. We presented basic test accuracy statistics across all studies, and we assessed test accuracy at the various follow-up periods and thresholds described in the included studies.

In our protocol, we detailed planned assessments of heterogeneity relating to age, case mix, clinical criteria for diagnosing dementia, technical features of the testing strategy, and other factors specific to the delayed verification analysis. These analyses were not possible with the data in this review.

Sensitivity analyses

In our protocol, we specified certain sensitivity analyses to explore the sensitivity of any summary accuracy estimates to aspects of study quality, such as nature of blinding and loss to follow-up, guided by the anchoring statements developed in our QUADAS-2 exercise. These analyses were not possible with the data in this review.

Due to the potential for bias, we pre-specified that case-control data were not included.

Assessment of reporting bias

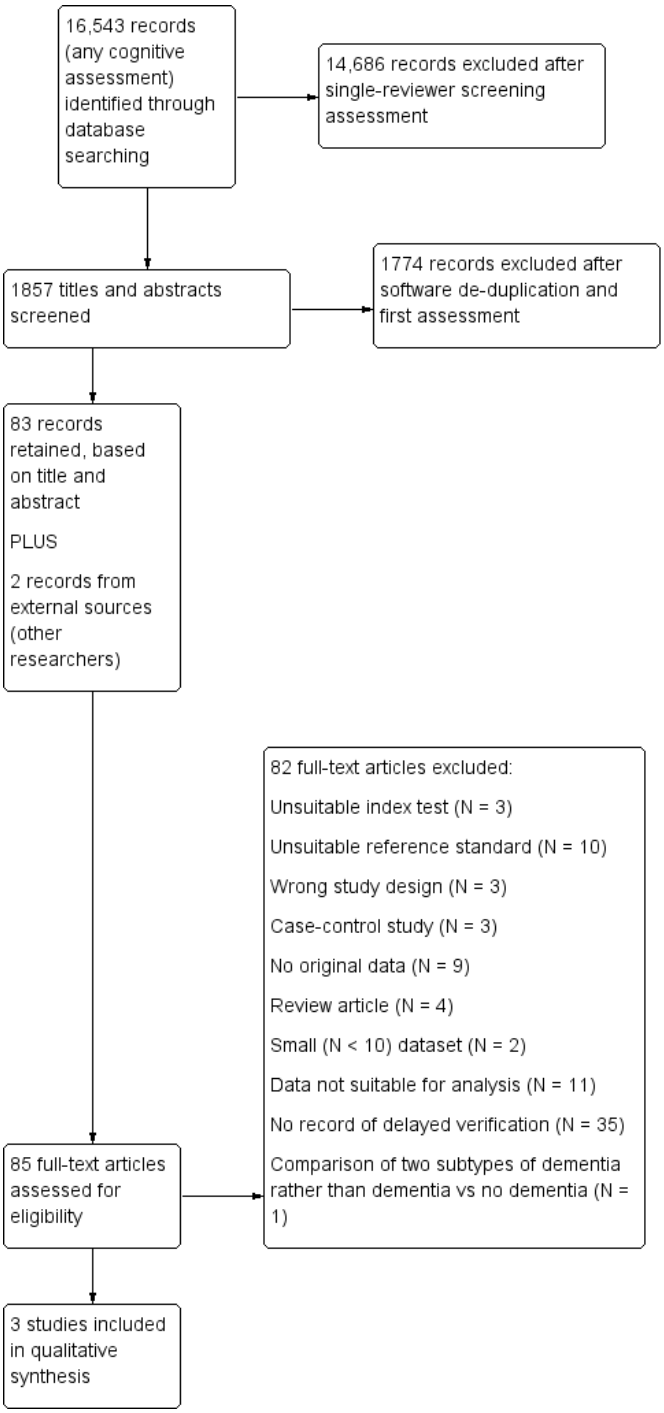
Reporting bias was not investigated because of current uncertainty about how it operates in test accuracy studies and in the interpretation of existing analytical tools, such as the funnel plot.

RESULTS

Results of the search

Our search identified 16,543 citations, from which we identified 85 full-text papers for potential eligibility. We excluded 82 papers (Figure 1). Reasons for exclusion were: no IQCODE data or unsuitable IQCODE data, small numbers (< 10) of included participants, no clinical diagnosis of dementia, repeat data sets, data not suitable for analysis (described in more detail in Selection of studies), no data regarding delayed verification, wrong study design, and case-control design (see Characteristics of excluded studies).

Figure 1. Study flow diagram.



Eight studies required translation. We contacted 19 authors to provide useable data, 16 of whom responded (see Acknowledgements).

This review includes three studies, N = 626 participants (Summary of findings 1). None of the included studies were described as primary delayed verification studies, and the original papers did not have an exclusive delayed verification accuracy focus. We obtained additional data from all three author groups in correspondence to facilitate inclusion in the review.

Methodological quality of included studies

We described the risk of bias using the QUADAS-2 methodology (Appendix 5), and we assessed reporting quality with STARDdem (Appendix 7); our anchoring statements for the IQCODE are summarised in Appendix 6. We did not rate any study as having low risk of bias for all the categories of QUADAS-2 (Figure 2; Figure 3).

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

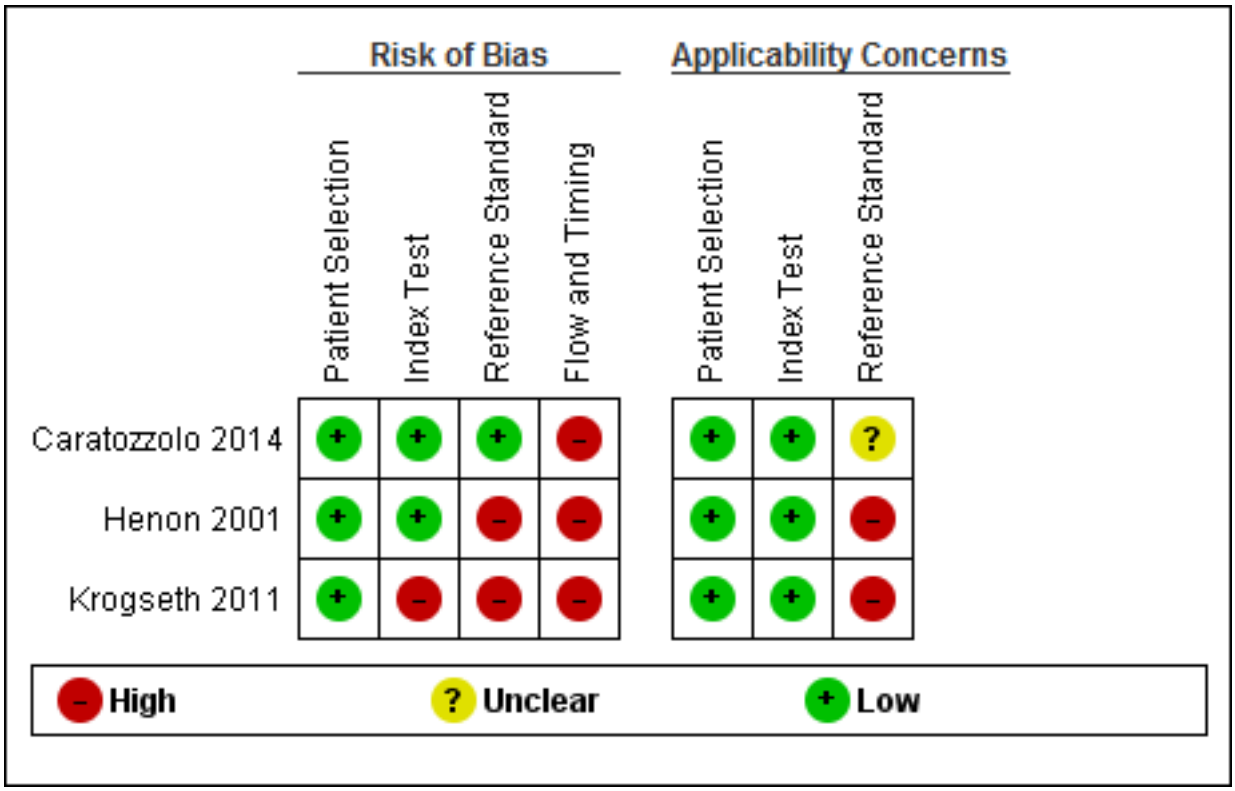
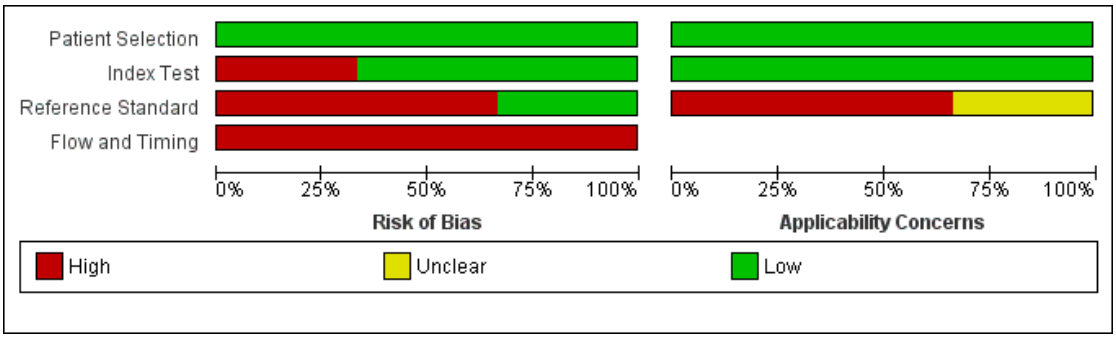


Figure 3. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies



Patient selection/sampling

All studies were at low risk of bias for patient selection, based on our pre-defined anchoring statements. All three were of cohort design and avoided inappropriate exclusions. One study sought a consecutive sample of admissions (Henon 2001). However, all studies excluded those who did not have an informant to complete the IQCODE assessment, and all excluded those who had pre-existing dementia, thus, none recruited a consecutive sample of admissions.

In all three studies, we felt there was low concern about the applicability of the findings to the populations under study. Two studies were conducted in the acute stroke unit setting (Caratozzolo 2014; Henon 2001), and the final one was conducted on admissions for acute hip fracture (Krogseth 2011), both of which were considered common in-patient secondary care populations. This grading does not suggest that results from these studies in specialist areas could be extrapolated to an unselected population of older adults.

All studies used a method of excluding prevalent dementia, 22% (N = 135/626) of the total were assessed to have pre-stroke or pre-fracture dementia. The methods for reaching this diagnosis varied. In Krogseth 2011, determination of pre-fracture dementia was based on a review of the patient's medical records, including prior cognitive testing, brain imaging, or both. This was combined with their IQCODE, MMSE, and Clock-Drawing Test scores (Agrell 1998), and presented to two specialists who determined if the individual met DSM-IV criteria for dementia. In Caratozzolo 2014, pre-stroke dementia was defined by having an existing diagnosis of dementia using DSM-IV criteria. In Henon 2001, pre-stroke dementia was defined as having an IQCODE score of 104 or greater, which equates to a score of 4.0.

IQCODE (index test) application

One study was considered to be at high risk of bias in index test application, as the threshold used to define test positivity was not pre-specified, and was based on the baseline characteristics of recruited participants (Krogseth 2011). In the other two studies, IQCODE positivity was pre-specified at higher than 3.3 (Caratozzolo 2014), and 3.0 (Henon 2001), respectively. This assessment was difficult to operationalise for our delayed verification focus, where there was no guidance on an appropriate IQCODE threshold.

For all three studies, there was low concern about the applicability of the conduct or interpretation of the index test.

Dementia diagnosis (reference standard) application

Two studies were at high risk of bias in the use of the reference standard (Henon 2001; Krogseth 2011). Henon 2001 reached a reference standard diagnosis in a diagnostic case conference forum. However, not all included participants received the same reference standard, and where participants were not assessed, the index test was used to determine the reference standard. Krogseth 2011 used the results of the index test to inform the creation of the reference standard diagnosis.

Caratozzolo 2014 was at low risk of bias in this domain, as the reference standard diagnosis was made by clinicians blinded to the results of the index test. However, the method for reference standard assessment was not described in the study abstract, and thus the applicability was graded as unclear. In subsequent correspondence with the author team, the method used was based on the Itel-MMSE (Metitieri 2001), with a score less than 24, the Barthel Index (Mahoney 1965), and an Instrumental Activities of Daily Living scale that indicated the loss of more than one activity of daily living. This defined states of 'possible post-stroke dementia' and 'no dementia'; these categories were then appraised by a neurologist using DSM-IV criteria. We felt the applicability of this two-stage process was uncertain, and the grading of unclear was maintained.

Flow and timing

There was substantial attrition. The three studies had a baseline population of N = 626, 47% (N = 295) of whom received the reference standard assessment at the first follow-up period, which ranged from three to six months, and 28% (N = 174) of whom received the reference standard assessment at the final follow-up period, which ranged from one to three years.

All three studies were at high risk of bias for the domain of flow and timing. The longitudinal nature of the studies resulted in significant attrition, either due to death or loss to follow-up. Missing data for participants were an issue in all three studies.

Reporting quality

Reporting quality tools exist for various study designs. STARDdem guidance is structured around key aspects of reporting that is required in test accuracy studies; reporting quality was described for each study using the STARDdem guidance (Appendix 4), which is presented in Appendix 7. Important limitations in reporting were the number, training, and expertise of the persons executing and reading the index tests and reference standard; blinding of the readers of the index test and reference standard, and how indeterminate results, missing data, and outliers of the index tests were handled.

Findings

The included study characteristics are described in the Characteristics of included studies, Summary of findings 1, and Summary of findings 2.

Caratozzolo 2014 recruited 121 acute stroke inpatients, free of dementia at baseline, and assessed them for the presence of dementia at three months and one year of follow-up. IQCODE data were available at baseline for all included participants, 114 were assessed at three months, and 105 at one year, with all losses due to death in the intervening period. The prevalence of dementia was 25% at three months, and 35% at one year.

Using a cut-off of higher than 3.3, the IQCODE had a sensitivity of 0.86 (95%CI 0.67 to 0.96) and a specificity of 0.90 (95%CI 0.81 to 0.95) for the clinical diagnosis of dementia at three months, and a sensitivity of 0.84 (95%CI 0.68 to 0.94) and a specificity of 0.87 (95%CI 0.76 to 0.94) for the clinical diagnosis of dementia at one year.

Henon 2001 recruited acute stroke inpatients, free of dementia at baseline, and assessed them for the presence of dementia at six months, one year, two years, and three years of follow-up. From an initial sample of 169 individuals, there was significant attrition at each follow-up period, due to patient death and unwillingness for further assessment. At six months, 99 participants were assessed, 85 were assessed at one year, 65 were assessed at two years, and 69 participants were assessed at three years. Around 25% of the participants had died by the six-month follow-up; this rose to 38% by the three-year follow-up. When individuals were not assessed by the study neurologist, the authors used additional means of evaluating dementia status, including telephone contact with the general practitioner or family members. Prevalence of dementia was 26% at six months, 24% at one year, 20% at two years, and 16% at three years.

Using a cut-off of higher than 3.0, the IQCODE had a sensitivity of 0.77 (95%CI 0.56 to 0.91) and a specificity of 0.51 (95%CI 0.39 to 0.63) for the clinical diagnosis of dementia at six months, and a sensitivity of 0.75 (95%CI 0.51 to 0.91) and a specificity of 0.46 (95%CI: 0.34 to 0.59) at one year. At two years, the sensitivity was 0.85 (95%CI 0.55 to 0.98) and specificity was 0.46 (95%CI: 0.32 to 0.61), and at three years, the sensitivity was 0.82 (95%CI 0.48 to 0.98) and specificity was 0.38 (95%CI 0.26 to 0.52) for the clinical diagnosis of dementia.

Krogseth 2011 recruited hip fracture inpatients and evaluated the effects of delirium on the risk of incident dementia at six-month follow-up. Data on the IQCODE assessment at baseline were missing for 25% (27/106) of included participants, leaving 82 who were assessed at baseline and at six months. Prevalence of dementia at follow-up was 12%.

Using a cut-off of higher than 3.12, the IQCODE had a sensitivity of 0.80 (95% CI 0.44 to 0.97) and specificity of 0.53 (95C%CI 0.41 to 0.65) for the clinical diagnosis of dementia at six months.

We did not perform meta-analyses to describe summery estimates of interest. In our protocol, we had pre-specified that more than three studies would be required for a meta-analysis to be valid. We were also mindful of the heterogeneity between the included studies, which described very different healthcare settings and patient populations. Had we found a larger number of studies, we could have pooled data and then investigated the effects of certain study characteristics on the accuracy of estimates, using meta-regression, however, with the modest number of studies in this review, such an analysis was not possible. In view of the heterogeneity between the three included studies, the lack of agreed threshold for IQCODE positivity, and lack of common follow-up, we were also unable to perform any of our pre-specified subgroup or sensitivity analyses.

Summary of findings

Study ID	Country	Subjects at Baseline (n)	Mean Age (yrs)	IQCODE Version	Language	Dementia Diagnosis	Dementia prevalence at 1st follow-up n/assessed (%) Timing	Dementia Prevalence at last follow-up n/assessed (%) Timing	Other Assessments
Caratozzolo 2014	Italy	158	68.4 to 77.4	16-item	Italian	DSM-IV	28/114 (24.6) 3 months	37/105 (35.2) 12 months	BI; IADL; Itel-MMSE
Henon 2001	France	202	≥ 40	26-item	French	ICD-10	26/99 (26.2) 6 months	11/69 (15.9) 3 years	MDRS, MADRS, MMSE
Krogseth 2011	Norway	266	82.7	16-Item	Norwegian	DSM-IV	10/82 (12.2) 6 months	*	CAM, MMSE, CDT, ADL

Abbreviations: ADL- Activities in Daily Living; BI- Barthel Index; CAM- Confusion Assessment Method; CDT- Clock Drawing Test; DSM- American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders; IADL- Instrumental Activities of Daily Living; ICD- International Classification of Disease; Itel-MMSE- Italian version of MMSE; MADRS- Montgomery-Asberg Depression Rating Scale; MDRS- Mattis Dementia Rating Scale; MMSE- Mini-Mental State Examination.

* only single time point of assessment

What is the accuracy of the Informant Questionnaire for Cognitve Decline in the Elderly (IQCODE) test for the early diagnosis of dementia when differing thresholds are used to define IQCODE positive cases?	
Population	Adults, free of dementia at baseline assessment, who were assessed using the IQCODE, some of who will develop dementia over a period of follow-up. The implication is that at the time of testing, the individual had a cognitive problem sufficient to be picked up on screening, but not yet meeting dementia diagnostic criteria
Setting	We considered any use of IQCODE as an initial assessment for cognitive decline, and we did not limit studies to a particular healthcare setting. We operationalised the various settings where the IQCODE may be used as secondary care, primary care, and community
Index test	Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE), administered to a relevant informant. We restricted analyses to the traditional 26-item IQCODE and the commonly-used short form IQCODE with 16 items
Ref**erence Standard**	Clinical diagnosis of dementia made using any recognised classification system
Studies	We included cross-sectional studies but not case-control studies

Test	Summary accuracy (95% CI)	No. of participants (timeframe)	Dementia prevalence	Implications, Quality and Comments
IQCODE cut-off 3.0	At six months: Sensitivity 0.77 (0.56 to 0.91) ; Specificity 0.51 (0.39 to 0.63) At one year: Sensitivity 0.75 (0.51 to 0.91) ; Specificity 0.46 (0.34 to 0.59) At two years: Sensitivity 0.85 (0.55 to 0.98) ; Specificity 0.46 (0.32 to 0.61) At three years: Sensitivity 0.82 (0.48 to 0.98) ; Specificity of 0.38 (0.26 to 0.52)	From 1 study: 99 (at 6 months) 85 (at 1 year) 65 (at 2 years) 69 (at 3 years)	26% (at 6 months) 24% (at 1 year) 20% (at 2 years) 16% (at 3 years)	Using three thresholds to define IQCODE test positivity, the IQCODE appeared to be relatively sensitive in diagnosing dementia at follow-up over 3 months to 3 years All included participants were hospitalised either for acute stroke or hip fracture. The findings could not be pooled and do not allow for recommendations for clinical practice
IQCODE cut-off 3.12	At six months: Sensitivity 0.80 (0.44 to 0.97) ; Specificity 0.53 (0.41 to 0.65)	From 1 study: 82 (at 6 months)	12% (at 6 months)	

IQCODE cut-off 3.3	At three months: Sensitivity 0.86 (0.67 to 0.96); Specificity 0.90 (0.81 to 0.95) At one year: Sensitivity 0.84 (0.68 to 0.94); Specificity 0.87 (0.76 to 0.94)	From 1 study: 114 (at 3 months) 105 (at 1 year)	25% (at 3 months) 35% (at 1 year)

DISCUSSION

Summary of main results

Our review identified three heterogeneous studies, with follow-up evaluation of dementia at time points between three months and three years. The included studies all reported on patients at high risk of developing a cognitive syndrome due to either delirium or stroke.

The IQCODE was used at three thresholds of positivity (higher than 3.0, higher than 3.12, and higher than 3.3) to predict those at risk of a future diagnosis of dementia. Using the higher than 3.3 threshold, Caratozzolo 2014 found a modest sensitivity with higher specificity for identifying those who would develop dementia at three months and one year of follow-up. For the lower thresholds of higher than 3.0 and higher than 3.12, used by Henon 2001 and Krogseth 2011 respectively, the IQCODE was again modestly sensitive, but lacked specificity. Test accuracy fell over time, with significant attrition of participants limiting the numbers available at follow-up, and the confidence intervals associated with the summary properties widening as a consequence.

Methods for excluding prevalent dementia at baseline were varied, and all had potential for bias. Defining pre-stroke dementia, based on a high IQCODE score, was not ideal for a study of IQCODE properties, albeit this was not the authors' main focus in this study (Henon 2001). Case-note review for a label of dementia was likely to miss a proportion with early dementia (Caratozzolo 2014). These approaches had the potential to bias the test accuracy results, as they may have falsely reduced or inflated the disease prevalence. The method of assessing for the reference standard was also varied, with Henon 2001 using indirect assessments, including general practitioner data and telephone follow-up. Although this method sought to reduce losses to follow-up by using proxy information, it had the potential to dilute the quality and certainty of the reference standard assessment, which may have lead to misclassification.

Strengths and weaknesses of the review

Strengths and weaknesses of the included studies

Our risk of assessment of internal and external validity, using the QUADAS-2 tool, identified issues across many aspects of study design and conduct. This reflected both the methodological challenges of conducting cognitive studies with prospective follow-up and the challenges for reviewers of applying a quality assessment tool that is better suited to classical cross-sectional test accuracy reports.

All three studies recruited from secondary care inpatient settings, two with an acute stroke focus (Caratozzolo 2014; Henon 2001), and the other describing cognition following hip fracture (Krogseth 2011). These were selected populations who had experienced physiological insult and brain injury (for the majority) and who were at high risk of subsequently developing dementia (Bejot 2011; Davis 2012). This would increase the prevalence of our reference standard at follow-up and so limited the generalisability of the findings to other non-acute settings. We did not identify any studies that evaluated the performance of the IQCODE in identifying those who would go on to develop dementia without the presence of an acute event at the time of assessment.

To align with the delayed verification focus, clarifying dementia status at baseline was fundamental to the study design. There is no guidance on the preferred strategy for retrospectively assessing dementia status following a major insult such as stroke or fracture (McGovern 2016). The definition of pre-stroke dementia used by Henon 2001, used the IQCODE in isolation and had more potential for bias than the clinical assessment method used by Krogseth 2011. Caratozzolo 2014 did not actively assess dementia at the time of first presentation, instead relying on individuals having an established diagnosis. This approach may have meant that individuals with undiagnosed dementia were included in the analysis, as it is known that dementia is under-diagnosed in those who present for acute hospital care (Sampson 2009).

The use of IQCODE varied across the studies. We note, in common with other IQCODE reviews, that availability of an informant was not guaranteed. This immediately created potential for bias as those with no available informant were likely to differ from those who had someone that could complete the IQCODE. The studies used IQCODE cut-offs that differed from those used to indicate probable dementia; this was appropriate, as the purpose of testing was not to diagnose contemporary dementia but to look at a future risk.

The choice of IQCODE cut-off used was interesting, with Henon 2001 using any score above 3.0 (where 3.0 indicated no change over the last ten years). This may explain the high sensitivity but poor specificity of the tool. There is no guidance on a suitable cut-off if using the IQCODE to assess future risk of dementia, but we would assume that the threshold used would be lower than that used to define dementia. The cut-off of 3.3 used in Caratozzolo 2014 has been used to define contemporaneous dementia in previous studies (Harrison 2014; Harrison 2015). Whether the initial IQCODE was assessing for a pre-dementia state or was assessing for early undiagnosed dementia is debatable. The follow-up periods (in months) used in some of the studies seemed rather short to allow for the development of incident dementia. The 'natural history' of cognitive change following stroke and fracture are not well described (Brainin 2015), and this further limited the interpretation of our results. There is no consensus on the optimal time point to assess for progression of dementia. Although our review did not have an MCI focus, the MCI literature suggests that it can take several years for a substantial proportion of patients to 'convert' to dementia (Ritchie 2015). The population of inter-

est in this review had a dementia syndrome, but at a very early stage. Even if this population progressed at a faster rate than MCI convertors, follow-up would still have to be in the order of years, rather than months. We pre-specified that we would assess for use of interventions that may impact on the usual cognitive trajectory. No studies gave this level of detail, but arguably, this was not an issue, since we currently have no evidence-based intervention that impacts meaningfully on cognitive decline.

The assessment of the reference standard, clinical diagnosis of dementia, also varied between studies. As with other reviews of IQCODE, we noted the possible biases from the incorporation of the index test (IQCODE) into the reference standard assessment. This bias may have been difficult to avoid, as our chosen reference standard, clinical assessment of dementia, is itself partly based on structured collateral history from an informant. The question around timing of assessment for our reference standard was equally challenging.

Although the follow-up was not particularly long, there was substantial attrition over time. This reflected the sampling frame; both stroke and fractured neck of femur are associated with short to medium-term mortality and institutionalisation. The loss to follow-up was unlikely to be random, and those at greatest risk of dementia were likely to be over represented in the population with no follow-up assessment. This explained the counterintuitive finding of decreasing prevalence of dementia over time in the study with the longest follow-up (Henon 2001). There is no consensus on how to deal with missing data in the context of competing risk for a delayed verification test accuracy design. However, this situation is likely to be common to other studies that look at the prospective development of dementia in an older adult cohort.

To allow a comprehensive assessment of the included studies, we complemented our QUADAS-2 review with an assessment of quality of reporting. We used a dementia-specific extension to STARD (STARDdem (Noel-Storr 2014)), but as our chosen papers were not framed as test accuracy studies per se, it was difficult to apply the STARDdem criteria. Accepting this caveat, our STARDdem assessment highlighted some limitations in reporting that seemed to be common to other dementia test accuracy studies. Lack of detail on how missing data, uninterpretable results, and losses to follow-up were accounted for in the papers was a concern, and we would urge greater detail and transparency around these issues for future studies.

Strengths and weaknesses of the review process

The review benefits from a robust search methodology applied to a targeted population. This identified only three studies suitable for inclusion, none of which were primarily designed as diagnostic test accuracy studies. We would argue that this finding reflects a lack of research in this area, rather than an overly focused search strategy, as an equivalent search identified substantial numbers of studies assessing IQCODE's use in secondary care (Harrison 2015), and community settings (Quinn 2014).

We operated no exclusions with regard to study language or year of publication. As part of the suite of reviews describing IQCODE, we have contacted research teams with an interest in cognitive screening to check for unpublished or in press original data. Where reporting was not clear in the included manuscript, we contacted the study authors, who supplied additional details; this enabled us to include data from all three of the studies in this review.

The review is strengthened by the application of formal, dementia-specific tools for the assessment of methodological and reporting quality. We used QUADAS-2-based anchoring statements specifically developed for use with studies that have a cognitive index test or reference standard (Davis 2013). Our complementary assessment of reporting used the dementia-focused extension to standard guidelines STARDdem (Noel-Storr 2014). Although these tools were the most appropriate for our study question, they were primarily developed for cross-sectional test accuracy work, and we experienced some difficulty in aligning them with the delayed verification approach.

The delayed verification research design is frequently used in studies of dementia biomarkers, particularly those biomarkers that purport to define a pre-clinical stage of disease. In designing our suite of test accuracy reviews for IQCODE, we included the delayed verification design. With hindsight, delayed verification is difficult to operationalise with questionnaire-based cognitive testing. The complexity increases when considering IQCODE, a tool that is based on symptoms over the preceding ten years. Thus, we were describing the use of a retrospective assessment for assigning potential prospective disease status.

Comparisons with previous research

This review forms part of a series of reviews describing informant-based cognitive screening tools. Other reviews describing IQCODE use in a primary care (Harrison 2014), community (Quinn 2014) or hospital context (Harrison 2015), are available. The heterogeneity of approaches used to define IQCODE positivity is in common with the previous reviews in the series.

We set a specific review question around IQCODE assessment in a population with no dementia. Other papers have used baseline IQCODE and prospective follow-up in different and perhaps more clinically meaningful ways. Jackson 2014, one of the studies excluded from this review, took an alternative approach to using the IQCODE as a tool for detecting dementia. This test accuracy study used the IQCODE at the time of acute hospital presentation for delirium and then re-evaluated individuals at three-month follow-up. This evaluation allowed for the exclusion of ongoing delirium and evaluation of the status of the individual following their acute admission, seeking to identify undiagnosed dementia. Using the IQCODE at a cut-off of higher than 3.65 offered the most favourable results (Jackson 2014).

Applicability of findings to the review question

The delayed verification model in test accuracy has been developed to evaluate any test that suggests it can identify those who have preclinical dementia. This area of research is dominated by the desire to identify and define biomarkers of early disease, matched with an understandable desire to identify targets for therapeutic intervention to prevent or delay disease progression. Intuitively, it should hold that neuropsychological assessments, both direct and informant-based, should identify such individuals, although data in this area have been very limited. This review identified some of the key challenges in conducting such studies, primarily attrition over time, although in both cases, acutely unwell hospitalised older adults were the subjects, who may be more prone to early mortality. As a tool for delayed verification, the IQCODE has potential limitations, and may not be suited to detecting pre-clinical disease. In the included papers, it is debatable what the IQCODE is detecting. Although the papers describe excluding prevalent dementia, the assessment of dementia was not robust in all the studies and it is likely that patients with early (undiagnosed) dementia were included and 'conversion' to dementia at follow-up simply represented progression of the underlying disease. The included papers did not exclude participants with baseline MCI, who were also likely to make up a proportion of the 'convertors' to dementia.
We specified a number of subgroup and sensitivity analyses of interest, but the limited data available precluded our progressing these. Questions remain around the potential differential properties of delayed verification when considering an insidious, progressive neurodegenerative process like Alzheimer's Disease dementia and major neurocognitive disorders that can have a more abrupt onset, such as vascular cognitive impairment.

AUTHORS' CONCLUSIONS

Implications for practice

The studies identified did not allow us to make specific recommendations on the use of the IQCODE for the early diagnosis of dementia in clinical practice. Indeed, it is debatable whether IQCODE is suited to this purpose. However, our review question was not irrelevant, as IQCODE is used in practice to predict future cognition in certain areas, such as acute stroke (McGovern 2016). If IQCODE is to be used in this way, the limited available data suggest that it is sensitive but not sufficiently specific to inform clinical decision-making. In this situation, clinicians may wish to complement the IQCODE with another more specific baseline assessment, or they may wish to adopt a two-stage screening, with initial IQCODE testing and then further testing of all 'positive' cases with a more specific tool.

Implications for research

The available evidence suggests that researching the IQCODE as a diagnostic tool for the delayed verification of dementia is challenging, with significant loss to follow-up over time affecting estimates of diagnostic accuracy. Future work must be explicit about this issue and how to deal with losses. This may require an assessment of the nature of reference standard assessment procedures, and whether comprehensive face-to-face assessment can be performed in all cases. The adequacy of alternative approaches, such as telephone assessment, would need to be established, given that the gold standard, clinical diagnosis of dementia, requires a multidimensional approach. An alternative approach may be the use of data linkage technology to ascertain diagnostic status over longitudinal follow-up. However, such approaches may be limited by the recording of dementia diagnosis on healthcare records and death certificates, which is known to be sub-optimal (Romero 2014), and the risk of missing those who have not yet received a formal diagnosis (Bamford 2007).

ACKNOWLEDGEMENTS

We thank the following researchers who assisted with translation: Salvador Fudio, EMM van de Kamp-van de Glind, Anja Hayen.

We thank the following researchers who responded to requests for original data: Dr S Caratozzolo, Dr JFM de Jonghe, Dr T Girard, Dr D Goncalves, Prof H Henon, Dr V Isella, Dr T Jackson, Dr M Krogseth, Dr AJ Larner, Dr K Okanurak, Dr G Potter, Dr M Razavi, Dr B Rovner, Dr D Salmon, Dr S Sikkes and Dr V Valcour.

We would like to thank Dr Yemisi Takwoingi for providing one-on-one training with two of the review authors (JKH and TJQ) to facilitate data analysis.

REFERENCES

References to studies included in this review

Caratozzolo 2014 *{published and unpublished data}*

Caratozzolo S, Mombelli G, Riva M, Zanetti M, Gottardi F, Rozzini L, et al. New onset of dementia or previous cognitive impairment: a 3-month and 1 year after stroke overview. Neuroepidemiology. 2014:159.

Caratozzolo S, Riva M, Chilovi BV, Cerea E, Mombelli G, Padovani A, et al. Prestroke Dementia: characteristics and clinical features in consecutive series of patients. *European Neurology* 2014;**71**:148–54.

Henon 2001 *{published and unpublished data}*

Henon H, Durieu I, Guerouaou D, Lebert F, Pasquier F, Leys D. Poststroke dementia: incidence and relationship to prestroke cognitive decline. *Neurology* 2001;**57**:1216–22.

Krogseth 2011 *{published and unpublished data}*

Krogseth M, Wyller TB, Engedal K, Juliebo V. Delirium is an important predictor of incident dementia among elderly hip fracture patients. *Dementia and Geriatric Cognitive Disorders* 2011;**31**:63–70.

References to studies excluded from this review

Abreu 2008 *{published data only}*

Abreu ID, Nunes PV, Diniz BS, Forlenza OV. Combining functional scales and cognitive tests in screening for mild cognitive impairment at a university based memory clinic in Brazil. *Revista Brasileira de Psiquiatria* 2008;**30**:346–9.

Blackburn 2013 *{published data only}*

Blackburn DJ, Fox L, Mangoyana M, Bath PMW. Repeat cognitive screening (MoCA, MMSE & ACE-R, IQCODE) in a high risk post-stroke population: data from the 'prevention of decline in cognition after stroke trial' (PODCAST). *Cerebrovascular diseases* 2013;**Supplement, proceedings of the 2013 European Stroke Conference**: s519.

Scutt P, Blackburn D, Krishnan K, Ballard C, Burns A, Ford GA, et al. Baseline characteristics, analysis plan and report on feasibility for the Prevention Of Decline in Cognition After Stroke Trial (PODCAST). *Trials* 2015;**16**:509.

Bloomfield 2012 *{published data only}*

Bloomfield K, John N. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) completion on an acute care ward for the elderly: A brief study of informant characteristics. *International Psychogeriatrics* 2012;**24**(10):1700–1.

Bosboom 2013 *{published data only}*

Bosboom PR, Alfonso H, Almeida OP. Determining the predictors of change in quality of life self-ratings and carer-ratings for community-dwelling people with Alzheimer disease. *Alzheimer Disease and Associated Disorders* 2013;**27**(4):363–71.

Burke 2014 *{published data only}*

ACP Journal Club. Review: several brief screening tests detect dementia in older adults; no trials assess effects on patient outcomes. *Annals of Internal Medicine* 2014;**160**(4): JC12.

Burton 2015 *{published data only}*

Burton L, Tyson SF. Screening for cognitive impairment after stroke: a systematic review of psychometric properties and clinical utility. *Journal of Rehabilitation Medicine* 2015;**47**(3):193–203.

Butt 2008 *{published data only}*

Butt Z. Sensitivity of the IQCODE an application of item response theory. *Aging, Neuropsychology and Cognition* 2008;**15**:642–55.

Bystad 2013 *{published data only}*

Bystad M, Skjerve A, Strobel C. Assessing dementia: a presentation of MMSE-NR, clock-drawing test and informant scales [Demensutredning ved hjelp av MMSE–NR, klokketesten og pårørendeskalaer]. *Tidsskrift for Norsk Psykologforening* 2013;**50**(1):7–11.

Cherbuin 2008 *{published data only}*

Cherbuin N, Anstey KJ, Lipnicki DM. Screening for dementia: a review of self and informant assessment instruments. *International Psychogeriatrics* 2008;**20**:431–58.

Cherbuin 2012 *{published data only}*

Cherbuin N, Jorm AF. The IQCODE: using informant reports to assess cognitive change in the clinic and in older adults living in the community. *Cognitive screening instruments: A practical approach*. Springer, 2012:165–82.

Cruz-Orduna 2012 *{published data only}*

Cruz-Orduna I, Bellon JM, Torrero P, Aparicio E, Sanz A, Mula N, et al. Detecting MCI and dementia in primary care: efficiency of the MMS, the FAQ and the IQCODE. *Family Practice* 2012;**29**(4):401–6.

de Jonge 1997 *{published data only}*

De Jonghe JFM. Differentiating between demented and psychiatric patients with the dutch version of IQCODE. *International Journal of Geriatric Psychiatry* 1997;**12**:462–5.

Dekkers 2009 *{published data only}*

Dekkers M, Joosten-Weyn Banningh EW, Eling PA. Awareness in patients with mild cognitive impairment (MCI). *Tijdschrift voor Gerontologie en Geriatrie* 2009;**40**: 17–23.

Diefeldt 2007b *{published data only}*

Diesfeldt HFA. Informant based measures may over estimate cognitive impairment in elderly patients. *International Journal of Geriatric Psychiatry* 2007;**22**:1166–70.

Ehrensperger 2010 *{published data only}*

Enrensperger MM, Berres M, Taylor KI, Monsch AU. Screening properties of the German IQCODE with a two year time frame in MCI and early Alzheimer's disease. *International Psychogeriatrics* 2010;**22**:91–100.

Eramudugolla 2013 *{published data only}*

Eramudugolla R, Cherbuin N, Easteal S, Jorm AF, Anstey KJ. Self-reported cognitive decline on the Informant

Questionnaire on Cognitive Decline in the Elderly is associated with dementia, instrumental activities of daily living and depression but not longitudinal cognitive change. *Dementia and Geriatric Cognitive Disorders* 2013;**34**(5-6): 282–91.

Farias 2002 ***{published data only}***
Farias ST, Mungas D, Reed B, Haan MN, Jagust WJ. Everyday imaging in relation to cognitive functioning and neuroimaging in community dwelling Hispanic and non Hispanic older adults. *Journal of the International Neuropsychological Society* 2004;**10**:342–54.

Finneli 2009 ***{published data only}***
Finelli L, Kunze U, Gautier A, Gomez-Mancilla B, Monsch A. Algorithms to retrospectively diagnose mild cognitive impairment and dementia in a longitudinal study of ageing and dementia. *Alzheimers and Dementia* 2009;**5**: supplement.

Fuh 1995 ***{published data only}***
Fuh JL, Teng EL, Lin KN, Larson EB, Wang SJ, Liu CY, et al. The Informant Questionnaire on Cognitive Dcline in the Elderly as a screening tool for dementia for a predominantly illiterate Chinese population. *Neurology* 1995;**45**:92–6.

Garcia 2002 ***{published data only}***
Forcano Garcia M, Perlado Ortiz de Pinedo F. Cognitive deterioration: use of the short version of the Informant Test (IQCODE) in the geriatrics consultations. *Revista Espanola de Geriatria y Gerontolgia* 2002;**37**:81–5.

Girard 2014 ***{published data only}***
Girard TD, Edwards KM, Self WH, Grijalva CG, Zhu Y, Williams DJ, et al. Long-term cognitive outcomes after hospitalization for community-acquired pneumonia (COGCAP). *American Journal of Respiratory and Critical Care Medicine* 2014;**189**:D108.

Goncalves 2011 ***{published data only}***
Goncalves DC, Arnold E, Appadurai K, Byrne GJ. Case finding in dementia: comparative utility of three brief instruments in the memory clinic setting. *International Psychogeriatrics* 2011;**23**:788–96.

Hancock 2009 ***{published data only}***
Hancock P, Larner AJ. Diagnostic utility of the IQCODE and its combination with ACE-R in a memory clinic based population. *International Psychogeriatrics* 2009;**21**:526–30.

Harwood 1997 ***{published data only}***
Harwood DMJ, Hope T, Jacoby R. Cognitive impairment in medical inpatients - screening for dementia: is history better than mental state. *Age and Ageing* 1997;**26**:31–5.

Hayden 2003 ***{published data only}***
Hayden KM, Khachaturian AS, Tschanz JT, Corcoran C, Nortond M, Breitner JCS. Characteristics of a two-stage screen for incident dementia. *Journal of Clinical Epidemiology* 2003;**56**:1038–45.

Hollands 2015 ***{published data only}***
Hollands S, Lim YY, Buckley R, Pietrzak RH, Snyder PJ, Ames D, et al. Amyloid-beta related memory decline is not associated with subjective or informant rated cognitive impairment in healthy adults. *Journal of Alzheimer's Disease* 2015;**43**(2):677–86.

Isella 2002 ***{published data only}***
Isella V, Villa ML, Frattola L, Appollonio I. Screening cognitive decline in dementia preliminary data on the Italian version of the IQCODE. *Neurological Sciences* 2002; **23**:s79–80.

Isella 2006 ***{published data only}***
Isella V, Villa L, Russo A, Regazzoni R, Ferrarese C, Appollonio IM. Discriminative and predictive power of an informant report in mild cognitive impairment. *Journal of Neurology, Neurosurgery and Psychiatry* 2006;**77**:166–71.

Jackson 2014 ***{published and unpublished data}***
Jackson TA, Sheehan B, MacLullich AM, Lord HM, Gladman JR. Diagnostic test accuracy of informant based tools to diagnose dementia in older hospital patients with delirium: a prospective cohort study. *Age and Ageing* 2016 [Epub ahead of print 2016 Apr 13];**45**(4):505–11. ; DOI: 10.1093/ageing/afw065]
Jackson TA, Sheehan B, MacLullich AM, Lord HM, Gladman JR. Does the Informant Questionnaire of Cognitive Decline in the Elderly (IQCODE) predict dementia in patients admitted with delirium?. *European Journal of Geriatric Medicine* 2014;**supplement**:S67.

Jorm 1988a ***{published data only}***
Jorm AF, Korten AE. Assessment of cognitive decline in the elderly by informant interview. *British Journal of Psychiatry* 1988;**152**:209-13.

Jorm 1989 ***{published data only}***
Jorm AF, Jacomb PA. The informant questionnaire on cognitive decline in the elderly (IQCODE) sociodemographic correlates; reliability; validity and some norms. *Psychological Medicine* 1989;**19**:1015–22.

Jorm 1991 ***{published data only}***
Jorm AF, Scott R, Cullen JS, MacKinnon AJ. Performance of the IQCODE as a screening test for dementia. *Psychological Medicine* 1991;**21**:785–90.

Jorm 1994 ***{published data only}***
Jorm AF. A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) : development and cross validation. *Psychological Mmedicine* 1994;**24**: 145–53.

Jorm 1996A ***{published data only}***
Jorm AF, Christensen H, Henderson AS, Jacomb PA, Korten AE, Mackinnon A. Informant ratings of cognitive decline of elderly people relationships to longitudinal change on cognitive tests. *Age and Ageing* 1996;**25**:125–9.

Jorm 1997 ***{published data only}***
Jorm AF. Methods of screening for dementia: a meta-analysis of studies comparing an informant interview with a brief cognitive test. *Alzheimer Disease and Associated Disorders* 1997;**11**:158–62.

Jorm 2000 *{published data only}*
Jorm AF, Christensen H, Korten AE, Jacomb PA, Henderson AS. Informant ratings of cognitive decline in old age. *Psychological Medicine* 2000;**30**:981–5.

Jorm 2000a *{published data only}*
Jorm AF, Christensen H, Henderson AS, Jacomb PA, Korten AE, Mackinnon A. Informant ratings of cognitive decline in old age: validation against change on cognitive tests over 7-8 years. *Psychological Medicine* 2000;**30**:981–5.

Jorm 2003 *{published data only}*
Jorm AF. The value of informant reports for assessment and prediction of dementia. *Journal of the American Geriatrics Association* 2003;**51**:881–2.

Jorm 2004 *{published data only}*
Jorm AF. The IQCODE: a review. *International Psychogeriatrics* 2004;**16**:275–93.

Kathriarachi 2001 *{published data only}*
Kathriarachchi ST, Sivayogan S, Jayaratna SD, Dharmasena SR. Comparison of three instruments used in the assessment of dementia in Sri Lanka. *Indian Journal of Psychiatry* 2005; **47**:109–12.

Khachaturian 2000 *{published data only}*
Khachaturian AS, Gallo JJ, Breitner JC. Performance characteristics of a two-stage dementia screen in a population sample. *Journal of Clinical Epidemiology* 2000;**53**:531–40.

Knaefelc 2003 *{published data only}*
Knafelc R, Giudice DL, Harrigan S, Cook R, Flicker L, Mackinnon A, et al. The combination of cognitive testing and an informant questionnaire in screening for dementia. *Age and Ageing* 2003;**32**:541–7.

Larner 2010 *{published data only}*
Larner AJ. Can IQCODE differentiate Alzheimer's disease from frontotemporal dementia. *Age and Ageing* 2010;**39**: 392–4.

Larner 2013 *{published data only}*
Cherbuin N, Jorm AF. Chapter 8. The Informant Questionnaire for Cognitive Decline in the Elderly. In: Larner AJ editor(s). *Cognitive Screening Instruments*. London: Springer-Verlag, 2013:166–79.

Law 1995 *{published data only}*
Law S, Wolfson C. Validation of a French version of an informant based questionnaire as a screening test for Alzheimer's disease. *The British Journal of Psychiatry* 1995; **167**:541–4.

Li 2012 *{published data only}*
Li F, Jia XF, Jia J. The Informant Questionnaire on Cognitive Decline in the Elderly individuals in screening mild cognitive impairment with or without functional impairment. *Journal of Geriatric Psychiatry and Neurology* 2012;**25**:227–32.

Lin 2013 *{published data only}*
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):601–12.

Louis 1999 *{published data only}*
Louis B, Harwood D, Hope T, Jacoby R. Can an informant questionnaire be used to predict the development of dementia in medical inpatients?. *International Journal of Geriatric Psychiatry* 1999;**14**:941–5.

Mackinnon 1998 *{published data only}*
Mackinnon A, Mulligan R. Combining cognitive testing and informant report to increase accuracy in screening for dementia. *American Journal of Psychiatry* 1998;**155**: 1529–35.

Mackinnon 2003 *{published data only}*
Mackinnon A, Khalilian A, Jorm AF, Korten AE, Christensen H, Mulligan R. Improving screening accuracy for dementia in a community sample by augmenting cognitive testing with informant report. *Journal of Clinical Epidemiology* 2003;**56**:358-66.

Mimori 2000 *{published data only}*
Mimori Y. Cognitive decline and detection of dementia among the Japanese population analysis with CASI and IQCODE. Nihon Ronen Igakkai Zasshi. 2000; Vol. supplement:s451.

Morales 1995 *{published data only}*
Morales JM, Gonzalez-Montalvo JI, Bermejo F, Del-Ser T. The screening of mild dementia with a shortened Spanish version of the Informant Questionnaire on Cognitive Decline in the Elderly. *Alzheimer Disease and Associated Disorders* 1995;**9**:105–1.

Morales 1997a *{published data only}*
Morales JM, Bermejo F, Romero M, Del-Ser T. Screening of dementia in community dwelling elderly through informant report. *International Journal of Geriatric Psychiatry* 1997;**12**: 808–16. ; these are data from independent rural cohort

Morales 1997b *{published data only}*
Morales JM, Bermejo F, Romero M, Del-Ser T. Screening of dementia in community dwelling elderly through informant report. *International Journal of Geriatric Psychiatry* 1997;**12**: 808–16.

Morales-Gonzalez 1992 *{published data only}*
Morales-Gonzalez JM, Gonzalez-Montalvo JI, Del Ser Quijano T, Bermejo Pareja F. Validation of the S-IQCODE: the Spanish version of the informant questionnaire on cognitive decline in the elderly. *Archivos de neurobiologia* 1992;**55**(6):262–6.

Mulligan 1996 *{published data only}*
Mulligan R, Mackinnon A, Jorm A, Giannakopoulos P, Michel J. A comparison of alternative methods of screening for dementia in clinical settings. *Archives of Neurology* 1996; **53**:532–6.

Narasimhalu 2008 *{published data only}*
Narasimhalu K, Lee J, Auchus AP, Chen CPLH. Improviong detection of dementia in Asian Patients with low education: combining the MMSE and the IQCODE. *Dementia and Geriatric Cognitive Disorders* 2008;**25**:17–22.

Ozel-kizel 2010 *{published data only}*
Ozel-Kizel ET, Turan ED, Yilmaz E, Cangoz B, Uluc S. Discriminant validity and reliability of the Turkish version

of the IQCODE. *Archives of Clinical Neuropsychology* 2010; **25**:139–45.

Peroco 2009 ***{published data only}***
Perroco TR, Zevallos Bustamente SE, del Pilar Q Moreno M, Hototian SR, Lopes MA, Azevedo D, et al. Performance of Brazilian long and short IQCODE on the screening of dementia in elderly people with low education. *International Psychogeriatrics* 2009;**21**:531–8.

Potter 2009 ***{published data only}***
Potter GG, Plassman BL, Burke JR, Kabeto MU, Langa KM, Llewellyn DJ, et al. Cognitive performance and informant reports in the diagnosis of cognitive impairment and dementia in African Americans and whites. *Alzheimer's and Dementia* 2009;**5**:445–53.

Razavi 2011 ***{published data only}***
Razavi M, Margrett J, Oakland A, Martin P. Comparison of two informant questionnaire screening tools for dementia. abstract. 2011.

Ritchie 1992 ***{published data only}***
Ritchie K, Fuhrer R. A comparative study of the performance of screening tests for senile dementia using receiver operating characteristics analysis. *Journal of Clinical Epidemiology* 1992;**45**:627–37.

Rodriguez-Molinero 2010 ***{published data only}***
Rodriguez-Molinero A, Lopez-Dieguez M, Medina IP, Tabuenca AI, de la Cruz JJ, Banegas JR. Cognitive assessment of elderly patients in the emergency department. *Revista Espanola de Geriatria y Gerontolgia* 2010;**45**:183–8.

Rovner 2012 ***{published data only}***
Rovner BW, Casten RJ, Arenson C, Salzman B, Kornsey EB. Racial differences in the recognition of cognitive dysfunction in older persons. *Alzheimer Disease and Associated Disorders* 2012;**26**:44–9.

Sanchez 2009 ***{published data only}***
dos Santos Sanchez MA, Lourenco RA. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): cross-cultural adaptation for use in Brazil [Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): adaptação transcultural para uso no Brasil]. *Cadernos de Saúde Pública, Rio de Janiero* 2009;**25**: 1455–65.

Schofield 2006 ***{published data only}***
Schofield PW. Discrepancies in cognitive history from patient and informant in relation to cognitive function. *Research and Practice in Alzheimer's Disease* 2006;**11**:328–31.

Senanorong 2001 ***{published data only}***
Senanarong V, Assavisaraporn S, Sivasiriyanonds N, Printarakul T, Jamjumrus S, Udompunthurunk S, et al. The IQCODE an alternative screening test for dementia for low educated Thai elderly. *Journal of the Medical Association of Thailand* 2001;**84**:648–55.

Sikkes 2010 ***{published data only}***
Sikkes SAM, Van den Berg MT, Knol DL, de-Lange-de Klerk ESM, Scheltens P, Uitdehaag BMJ, et al. How useful is IQCODE for discriminating between Alzheimer's disease, mild cognitive impairment and subjective memory complaints?. *Dementia and Geriatric Cognitive Disorders* 2010;**30**:411–6.

Siri 2006 ***{published data only}***
Siri S, Okanurak K, Chansirikanjana S, Kitiyaporn D, Jorm AF. Modified IQCODE as a screening test for dementia for Thai elderly. *Southeast Asian Journal of Tropical Medicine and Public Health* 2006;**37**:587–94.

Srikanth 2006 ***{published data only}***
Srikanth V, Thrift AG, Fryer JL, Saling MM. The validity of brief screening cognitive assessments in the diagnosis of cognitive impairments and dementia after first ever stroke. *International Psychogeriatrics* 2006;**18**(2):295–305.

Starr 2000 ***{published data only}***
Starr JM, Nicolson C, Anderson K, Dennis MS, Deary IJ. Correlates of informant rated cognitive decline after stroke. *Cerebrovsacular Diseases* 2000;**10**(3):214–20.

Tang 2003 ***{published data only}***
Tang WK, Sandra SMC, Chiu HFK, Wong KS, Kwok TCY, Mok V, et al. Can IQCODE detect post stroke dementia? . *International Journal of Geriatric Psychiatry* 2003;**18**: 706–10.

Thomas 1994 ***{published data only}***
Thomas LD, Gonzales MF, Chamberlain A, Beyreuther K, Masters CL, Flicker L. Comparison of clinical state retrospective informant interview and the neuropathological diagnosis of Alzheimer's disease. *International Journal of Geriatric Psychiatry* 1994;**9**:233–6.

Tokuhara 2006 ***{published data only}***
Tokuhara KG, Valcour VG, Masaki KH, Blanchette PI. 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):72–5.

Wierderholt 1999 ***{published data only}***
Wiederholt WC, Galasko D, Salmon DP. Utility of CASI and IQCODE as screening instruments for dementia in natives of Guam. Journal of the Neurological Sciences. 1997; Vol. 150:s89.

Wolf 2009 ***{published data only}***
Wolf SA, Kubatschek K, Henry M, Harth S, Edbert AD, Wallesch CW. Informant report of cognitive changes in the elderly: a first evaluation of the German version of the IQCODE. *Nervenartz* 2009;**80**(10):1178–80.

Yamada 2000 ***{published data only}***
Yamada M, Mimori Y, Sasaki H, Ikeda J, Nakamura S, Kodama K. Cognitive dysfunction among the elderly evaluated by the cognitive abilities screening instrument. *Nihon Ronen Iqakkai Zasshi* 2000;**37**:56–62.

Zevallos-Bustamente 2003 ***{published data only}***
Zevallos Bustamante SE, Bottino CMC, Lopes MA, Dionísio Azevedo D, Hototatian SR, Litvoc J, et al. Combined instruments on the evaluation of dementia in the elderly: preliminary results. *Arquivos de Neuro-Psiquiatria* 2003;**61**:601–6.

Zhang 2003 *{published data only}*
Zhang XQ, Zhou JS, Wang LD, Meng C, Chen B. Memory complaints in the clinical diagnosis of dementia. *Chinese Journal of Clinical Rehabilitation* 2003;**7**:4254–5.

Zhou 2002 *{published data only}*
Zhou JS, Zhang XQ, Wang L. Telephone questionnaire: a new method for screening dementia. *Chinese Journal of Clinical Rehabilitation* 2002;**6**:3166–7.

Zhou 2003 *{published data only}*
Zhou J, Xinqing Z, Wang L, Meng C, Chu C, Chen B. Orientation memory concentration test and short IQCODE in the elderly: screen dementia by telephone. *Chinese Journal of Clinical Rehabilitation* 2003;**7**:1529–31.

Zhou 2004 *{published data only}*
Zhou JS, Zhang XQ, Mundt JC, Wang L, Meng C, Chu C, et al. Comparison of three dementia screening instruments administered by telephone in China. *The International Journal of Social Research and Practice* 2004;**3**:69–81.

Additional references

Agrell 1998
Agrell B, Dehlin O. The clock-drawing test. *Age and Ageing* 1998;**27**:399–403.

Bamford 2007
Bamford C, Eccles M, Steen N, Robinson L. Can primary care record review facilitate earlier diagnosis of dementia?. *Family Practice* 2007;**24**(2):108–16.

Bejot 2011
Bejot Y, Aboa-Eboule C, Durier J, Rouaud O, Jacquin A, Ponavoy E, et al. Prevalence of early dementia after first-ever stroke: a 24-year population-based study. *Stroke* 2011; **42**:607–12.

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 US Preventative Services Task Force. *Annals of Internal Medicine* 2003;**138**(11):927–37.

Brainin 2015
Brainin M, Tuomilehto J, Heiss WD, Bornstein N, Bath PM, Teuschl Y, et al. Post-stroke cognitive decline: an update and perspectives for clinical research. *European Journal of Neurology* 2015;**22**(2):229–38.

Brodaty 2002
Brodaty H. The GPCOG: a new screening test for dementia designed for general practice. *Journal of the American Geriatrics Society* 2002;**50**(3):530–4.

Chodosh 2004
Chodosh J, Petitti DB, Elliott M, Hays RD, Crooks VC, Reuben DB, et al. Physician recognition of cognitive impairment: evaluating the need for improvement. *Journal of the American Geriatrics Society* 2004;**52**(7):1051–9.

Cordell 2013
Cordell CB, Borson S, Boustani M, Chodosh J, Reuben D, Verghese J, et al. the Medicare Detection of the Cognitive Impairment Workgroup. Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during Medicare annual wellness visit in a primary care setting. *Alzheimer's & Dementia* 2013;**9** (2):141–50.

Davis 2012
Davis DH, Muniz Terrera G, Keage H, Rahkonen T, Oinas M, Matthews FE, et al. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. *Brain* 2012;**135**:2809–16.

Davis 2013
Davis DHJ, Creavin ST, Noel-Storr AH, Quinn TJ, Smailagic N, Hyde C, et al. Neuropsychological tests for the diagnosis of Alzheimer's disease dementia and other dementias: a generic protocol for cross-sectional and delayed-verification studies. *Cochrane Database of Systematic Reviews* 2013, Issue 3. [DOI: 10.1002/14651858.CD010460]

Ferri 2005
Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, et al. for Alzheimer's Disease International. Global prevalence of dementia: a Delphi consensus study. *Lancet* 2005;**366**(9503):2112–7.

Folstein 1975
Folstein MF, Folstein SE, McHugh PR. Minimental state: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;**12**(3): 189–98.

Galvin 2005
Galvin JE. The AD8: a brief informant interview to detect dementia. *Neurology* 2005;**65**(4):559–64.

Glanville 2012
Glanville J, Cikalo M, Crawford F, Dozier M, McIntosh H. Handsearching did not yield additional FDG-PET diagnostic test accuracy studies compared with electronic searches: a preliminary investigation. *Research Synthesis Methods* 2012;**3**:202–13.

Greenhalgh 1997
Greenhalgh T. Papers that report diagnostic or screening tests. *BMJ* 1997;**315**:540–3.

Harrison 2014
Harrison JK, Fearon P, Noel-Storr A-H, 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. [DOI: 10.1002/14651858.CD010771.pub2]

Harrison 2015
Harrison JK, Fearon P, Noel-Storr A-H, 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. [DOI: 10.1002/14651858.CD010772.pub2]

Herbert 2013
Herbert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* 2013;**80**(19):1778–83.

Jorm 1988
Jorm AF, Korten AE. Assessment of cognitive decline in the elderly by informant interview. *British Journal of Psychiatry* 1988;**152**:209–13.

Mahoney 1965
Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Medical Journal* 1965;**14**: 61–5.

McGovern 2016
McGovern A, Pendlebury ST, Mishra NK, Fan Y, Quinn TJ. Test accuracy of informant-based cognitive screening tests for diagnosis of dementia and multidomain cognitive impairment in stroke. *Stroke* 2016;**47**(2):329–35.

McKeith 2005
McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* 2005;**65**(12):1863–72.

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**(7): 939–44.

McKhann 2001
McKhann GM, Albert MS, Grossman M, Miller B, Dickson D, Trojanowski JQ, Work Group on Frontotemporal dementia and PIck's disease. Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. *Archives of Neurology* 2001;**58**(11):1803–9.

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 and the Alzheimer's Association workgroup. *Alzheimer's & Dementia* 2011;**7**(3):263–9.

Menon 2011
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**:272–6.

Metitieri 2001
Metitieri T, Geroldi C, Pezzini A, Frisoni GB, Bianchetti A, Trabucchi M. The Itel-MMSE: an Italian telephone version of the Mini-Mental State Examination. *International Journal of Geriatric Psychiatry* 2001;**16**(2):166–7.

Noel-Storr 2014
Noel-Storr AH, McCleery JM, Richard E, Ritchie CW, Flicker L, Cullum SJ, et al. Reporting standards for studies of diagnostic test accuracy in dementia: The STARDdem Initiative. *Neurology* 2014;**83**(4):364–73.

Peterson 2004
Petersen RC. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine* 2004;**256**:183–94.

Prince 2013
Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and meta-analysis. *Alzheimer's & Dementia* 2013;**9**(1): 63–75.

Quinn 2014
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, Issue 4. [DOI: 10.1002/ 14651858.CD010079.pub2]

RevMan 2014 [Computer program]
The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan) 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Ritchie 2015
Ritchie CW, Terrera GM, Quinn TJ. Dementia trials and dementia tribulations: methodological and analytical challenges in dementia research. *Alzheimer's Research & Therapy* 2015;**7**:31.

Roman 1993
Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. for the Vascular Dementia: Diagnostic Criteria for Research Studies work group. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology* 1993;**43**(2):250–60.

Romero 2014
Romero JP, Benito-Leon J, Mitchell AJ, Trincado R, Bermejo-Pareja F. Under reporting of dementia deaths on death certificates: using data from a population-based study (NEDICES). *Journal of Alzheimer's Disease* 2014;**39**(4): 741–8.

Sampson 2009
Sampson EL, Blanchard MR, Jones L, Tookman A, King M. Dementia in the acute hospital: prospective cohort study of prevalence and mortality. *British Journal of Psychiatry* 2009; **195**(1):61–6.

Valcour 2000
Valcour VG, Masaki KH, Curb JD, Blanchette PL. The detection of dementia in the primary care setting. *Archives of Internal Medicine* 2000;**160**(19):2964–8.

Wilson 1968
Wilson JMG, Jungner G, World Health Organization. Principles and practice of screening for disease. Geneva: World Health Organization 1968.

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Caratozzolo 2014

Study characteristics				**Study character**
Patient sampling	Inpatients hospitalised for acute stroke in an Italian hospital over an 8-month period			
Patient characteristics and setting	All inpatients admitted to the neurological clinic of an Italian hospital for a suspected acute cerebrovascular event were eligible for inclusion Participants were those who had: experienced an acute stroke, a diagnosis made by neurologists based on clinical symptoms and neuroimaging, and the availability of a reliable caregiver for each patient. Individuals known to have a diagnosis of dementia were excluded as were those experiencing a transient Ischaemic attack			
Index tests	IQCODE, 16-item, Italian language			
Target condition and reference standard(s)	Clinical diagnosis of dementia using DSM-IV criteria, diagnosed by a neurologist Assessment of reference standard not described in the abstract or original paper. In correspondence with authors, the assessment was conducted blinded to results of IQCODE			
Flow and timing	A total of 222 patients were evaluated, 64 of whom were excluded as they fulfilled the exclusion criteria or did not agree to participate in the study 158 were entered into the study, 37 of whom were diagnosed with having pre-stroke dementia and excluded, leaving 121 participants in the study At three months, 114 were assessed (five died during hospitalisation and two died during the follow-up period), and at one year, 105 were assessed (nine died between three- and twelve-month follow-up)			
Comparative				
Notes				
Methodological quality				**Methodological**
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**	
DOMAIN 1: Patient Selection				**DOMAIN 1: Pa**
Was a consecutive or random sample of patients enrolled?	No			
Was a case-control design avoided?	Yes			
Did the study avoid inappropriate exclusions?	Yes			

Caratozzolo 2014 (*Continued*)

		Low	**Low**
DOMAIN 2: Index Test All tests			**DOMAIN 2: In**
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			**DOMAIN 3: Re**
Is the reference standards likely to correctly classify the target condition?	Yes		
		Low	**Unclear**
DOMAIN 4: Flow and Timing			**DOMAIN 4: Fl**
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Henon 2001

Study characteristics		**Study character**
Patient sampling	Consecutive sample admitted to the stroke unit, excluding those with pre-stroke dementia (defined as IQCODE > 104), over a 28-week period	
Patient characteristics and setting	169 patients admitted to the acute stroke unit of a French university hospital. Participants were those experiencing an acute stroke, aged > 40 years, Caucasian, fluent French speakers who resided in the Lille community	
Index tests	IQCODE 26-item, French language	
Target condition and reference standard(s)	Clinical dementia diagnosis using ICD-10, applied at a diagnostic case conference that included the assessing neurologist and two specialist neuropsychologists, using data from neuropsychological testing or information from family or general practitioner, where formal testing was not possible, including a further IQCODE assessment	

Henon 2001 (*Continued*)

Flow and timing	258 potentially eligible patients at baseline, 56 of whom were excluded due to lack of informant or informant availability within 48 hours of stroke admission and 33 excluded due to the presence of pre-stroke dementia Follow-up was either a neurologist visit or telephone contact with the patient's family or the patient's general practitioner. Follow-up intervals were at 6 months, 1 year, 2 years, and 3 years post-event 65 died before the initial follow-up visit at 6 months and there was ongoing loss of participants; 127 at 6 months, 117 at 1 year, 111 at 2 years, and 104 at 3 years Not all recruited participants were prepared to be evaluated at follow-up; where this was the case, or they had died in the interval, information was obtained from their general practitioner or family, including an IQCODE assessment
Comparative	
Notes	

Methodological quality				**Methodological**
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**	
DOMAIN 1: Patient Selection				**DOMAIN 1: Pa**
Was a consecutive or random sample of patients enrolled?	Yes			
Was a case-control design avoided?	Yes			
Did the study avoid inappropriate exclusions?	Yes			
		Low	**Low**	
DOMAIN 2: Index Test All tests				**DOMAIN 2: In**
If a threshold was used, was it pre-specified?	Yes			
		Low	**Low**	
DOMAIN 3: Reference Standard				**DOMAIN 3: Re**
Is the reference standards likely to correctly classify the target condition?	No			
		High	**High**	
DOMAIN 4: Flow and Timing				**DOMAIN 4: Fl**

Henon 2001 (*Continued*)

Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Krogseth 2011

Study characteristics			**Study character**
Patient sampling	Hip fracture patients admitted to two Norwegian hospitals over a one-year period		
Patient characteristics and setting	106 patients who were admitted acutely with a hip fracture and operated on in two Norwegian hospitals. Eligibility was based on being over 65 years old, able to speak Norwegian, and length of stay > 48 hours. Exclusions were made for those with severe aphasia, head trauma, terminal illness, and prior inclusion in the study		
Index tests	IQCODE-16 item, Norwegian		
Target condition and reference standard(s)	An assessment of clinical diagnosis of dementia using DSM-IV criteria was made at two points during the study, at baseline and at six-month follow-up. The diagnosis was made by two study clinicians (one specialist in geriatric medicine and one specialist in geriatric psychiatry) At baseline, data were extracted from the participants' medical records for evidence of previous cognitive testing, hypothyroidism and B12 deficiency, and brain imaging. These data were combined with admission MMSE and CDT results and the pre-fracture IQCODE from their caregiver At 6-month follow-up, diagnosis was made using the results of cognitive testing, informant information about change in cognitive function post-fracture, and the report of the assessing physician. The assessing physician had made home visits for all included participants, conducting structured interviews and comprehensive cognitive testing		
Flow and timing	266 eligible patients, 92 of whom were lost to follow-up at six months (47 died, 35 declined, 2 moved, and 8 were participating in competing study). A further 65 were excluded as they were diagnosed with pre-fracture dementia, and 3 in whom pre-fracture cognition could not be assessed, leaving 106 participants with assessment at six months Not all of the included participants had available data for a baseline IQCODE (index test assessment) , 27 were missing		
Comparative			
Notes			

Methodological quality			**Methodological**
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Krogseth 2011 (*Continued*)

DOMAIN 1: Patient Selection				**DOMAIN 1: Pa**
Was a consecutive or random sample of patients enrolled?	Unclear			
Was a case-control design avoided?	Yes			
Did the study avoid inappropriate exclusions?	Yes			
		Low	**Low**	
DOMAIN 2: Index Test All tests				**DOMAIN 2: In**
If a threshold was used, was it pre-specified?	No			
		High	**Low**	
DOMAIN 3: Reference Standard				**DOMAIN 3: Re**
Is the reference standards likely to correctly classify the target condition?	Yes			
		High	**High**	
DOMAIN 4: Flow and Timing				**DOMAIN 4: Fl**
Was there an appropriate interval between index test and reference standard?	Yes			
Were all patients included in the analysis?	No			
		High		

IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly
DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
ICD-10 = International Statistical Classification of Diseases and Related Health Problems, 10th revision
MMSE = Mini-mental State Examination
CDT = Clock Drawing Test

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abreu 2008	not delayed verification
Blackburn 2013	data not suitable for analysis
Bloomfield 2012	wrong study design
Bosboom 2013	wrong study design
Burke 2014	no original data
Burton 2015	review article
Butt 2008	data on fewer than 10 subjects
Bystad 2013	review article
Cherbuin 2008	no new data
Cherbuin 2012	no original data
Cruz-Orduna 2012	no delayed verification
de Jonge 1997	data not suitable for analysis
Dekkers 2009	data not suitable for analysis
Diefeldt 2007b	no new data
Ehrensperger 2010	data not suitable for analysis
Eramudugolla 2013	wrong study design
Farias 2002	unsuitable reference standard
Finneli 2009	data not suitable for analysis
Fuh 1995	case-control study
Garcia 2002	not delayed verification
Girard 2014	unsuitable reference standard
Goncalves 2011	not delayed verification
Hancock 2009	not delayed verification

(*Continued*)

Harwood 1997	not delayed verification
Hayden 2003	data on fewer than 10 subjects
Hollands 2015	unsuitable reference standard
Isella 2002	data not suitable for analysis
Isella 2006	case-control design
Jackson 2014	no delayed verification
Jorm 1988a	not delayed verification
Jorm 1989	data not suitable for analysis
Jorm 1991	not delayed verification
Jorm 1994	not delayed verification
Jorm 1996A	unsuitable reference standard
Jorm 1997	no new data
Jorm 2000	unsuitable reference standard
Jorm 2000a	not delayed verification
Jorm 2003	no new data
Jorm 2004	no new data
Kathriarachi 2001	not delayed verification
Khachaturian 2000	data not suitable for analysis
Knaefelc 2003	not delayed verification
Larner 2010	two types of dementia rather than dementia versus no dementia
Larner 2013	review article
Law 1995	not delayed verification
Li 2012	unsuitable reference standard
Lin 2013	review article

(*Continued*)

Louis 1999	case-control design
Mackinnon 1998	not delayed verification
Mackinnon 2003	not delayed verification
Mimori 2000	no new data
Morales 1995	not delayed verification
Morales 1997a	not delayed verification
Morales 1997b	not delayed verification
Morales-Gonzalez 1992	not delayed verification
Mulligan 1996	not delayed verification
Narasimhalu 2008	not delayed verification
Ozel-kizel 2010	not delayed verification
Peroco 2009	not delayed verification
Potter 2009	data not suitable for analysis
Razavi 2011	not delayed verification
Ritchie 1992	data not suitable for analysis
Rodriguez-Molinero 2010	unsuitable reference standard
Rovner 2012	data not suitable for analysis
Sanchez 2009	unsuitable reference standard
Schofield 2006	data not suitable for analysis
Senanorong 2001	not delayed verification
Sikkes 2010	not delayed verification
Siri 2006	not delayed verification
Srikanth 2006	not delayed verification
Starr 2000	unsuitable reference standard

(*Continued*)

Tang 2003	not delayed verification
Thomas 1994	not delayed verification
Tokuhara 2006	not delayed verification
Wierderholt 1999	data not suitable for analysis
Wolf 2009	unsuitable reference standard
Yamada 2000	not delayed verification
Zevallos-Bustamente 2003	not delayed verification
Zhang 2003	data not suitable for analysis
Zhou 2002	not delayed verification
Zhou 2003	no new data
Zhou 2004	no new data

DATA

Presented below are all the data for all of the tests entered into the review.

Tests. Data tables by test

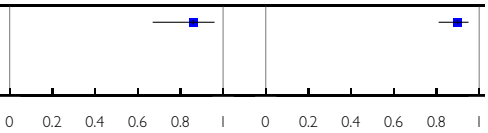
Test	No. of studies	No. of participants
1 IQCODE any threshold prediction of dementia at three months	1	114
2 IQCODE any threshold prediction of dementia at 6 months	2	181
3 IQCODE any threshold prediction dementia at 12 months	2	190
4 IQCODE any threshold prediction of dementia at two years	1	65
5 IQCODE any threshold prediction on dementia at three years	1	69

Test 1. IQCODE any threshold prediction of dementia at three months.

Review: Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings

Test: 1 IQCODE any threshold prediction of dementia at three months

Study	TP	FP	FN	TN	Sensitivity	Specificity	Sensitivity	Specificity
Caratozzolo 2014	24	9	4	77	0.86 [0.67, 0.96]	0.90 [0.81, 0.95]		

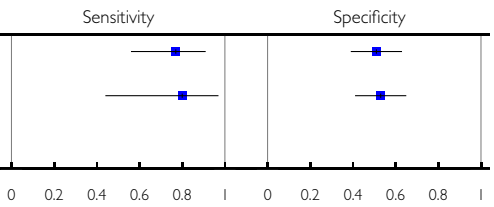


Test 2. IQCODE any threshold prediction of dementia at 6 months.

Review: Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings

Test: 2 IQCODE any threshold prediction of dementia at 6 months

Study	TP	FP	FN	TN	Sensitivity	Specificity
Henon 2001	20	36	6	37	0.77 [0.56, 0.91]	0.51 [0.39, 0.63]
Krogseth 2011	8	34	2	38	0.80 [0.44, 0.97]	0.53 [0.41, 0.65]

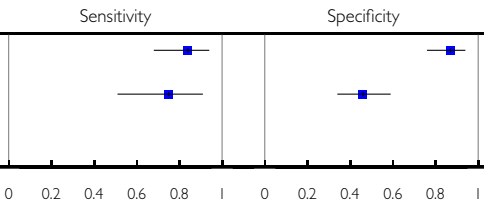


Test 3. IQCODE any threshold prediction dementia at 12 months.

Review: Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings

Test: 3 IQCODE any threshold prediction dementia at 12 months

Study	TP	FP	FN	TN	Sensitivity	Specificity
Caratozzolo 2014	31	9	6	59	0.84 [0.68, 0.94]	0.87 [0.76, 0.94]
Henon 2001	15	35	5	30	0.75 [0.51, 0.91]	0.46 [0.34, 0.59]

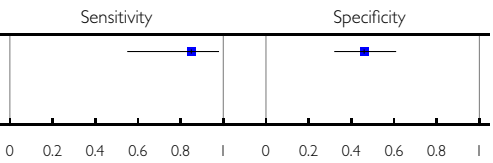


Test 4. IQCODE any threshold prediction of dementia at two years.

Review: Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings

Test: 4 IQCODE any threshold prediction of dementia at two years

Study	TP	FP	FN	TN	Sensitivity	Specificity
Henon 2001	11	28	2	24	0.85 [0.55, 0.98]	0.46 [0.32, 0.61]

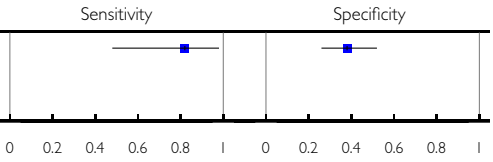


Test 5. IQCODE any threshold prediction on dementia at three years.

Review: Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings

Test: 5 IQCODE any threshold prediction on dementia at three years

Study	TP	FP	FN	TN	Sensitivity	Specificity
Henon 2001	9	36	2	22	0.82 [0.48, 0.98]	0.38 [0.26, 0.52]



APPENDICES

Appendix 1. Commonly used cognitive assessment or screening tools

TEST	Cochrane DTA review published/in progress
Mini-mental state examination (MMSE)	YES
GPcog	YES

(*Continued*)

Minicog	YES
Memory Impairment Screen (MIS)	Still available
Abbreviated mental testing	Still available
Clock-drawing tests (CDT)	Still available
Montreal Cognitive Assessment (MoCA)	YES
IQCODE (informant interview)	YES
AD-8 (informant interview)	YES

For each test, the planned review will encompass diagnostic test accuracy in community; primary and secondary care settings. As well as standard diagnosis, where applicable reviews will also describe delayed verification design trials.

Appendix 2. Search strategies

Source	Search strategy	Hits retrieved
1. MEDLINE In-process and other non-indexed citations and MEDLINE Ovid SP (1950 to 16 January 2016)	1. IQCODE.ti,ab. 2. "informant questionnaire on cognitive decline in the elderly".ti,ab 3. "IQ code".ti,ab. 4. ("informant* questionnair*" adj3 (dement* or screening)).ti,ab 5. ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. 6. or/1-5	Apr 2011: 291 Jul 2012: 39 Jan 2013: 19 Jan 2016: 46
2. Embase Ovid SP 1980 to 16 January 2016	1. IQCODE.ti,ab. 2. "informant questionnaire on cognitive decline in the elderly".ti,ab 3. "IQ code".ti,ab. 4. ("informant* questionnair*" adj3 (dement* or screening)).ti,ab 5. ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. 6. or/1-5	Apr 2011: 356 Jul 2012: 49 Jan 2013: 44 Jan 2016: 166
3. PsycINFO Ovid SP 1806 to January week 2 2016	1. IQCODE.ti,ab. 2. "informant questionnaire on cognitive decline in the elderly".ti,ab 3. "IQ code".ti,ab. 4. ("informant* questionnair*" adj3 (de-	Apr 2011: 215 Jul 2012: 28 Jan 2013: 17 Jan 2016: 50

(*Continued*)

	ment* or screening)).ti,ab 5. ("screening test*" adj2 (dement* or alzheimer*)).ti,ab. 6. or/1-5	
4. BIOSIS Previews (Thomson Reuters Web of Science) 1926 to 15 January 2016	Topic=(IQCODE OR "informant questionnaire on cognitive decline in the elderly" OR "IQ code") AND Topic=(dement* OR alzheimer* OR FTLD OR FTD OR "primary progressive aphasia" OR "progressive non-fluent aphasia" OR "frontotemporal lobar degeneration" OR "frontolobar degeneration" OR "frontal lobar degeneration" OR "pick* disease" OR "lewy bod*") Timespan=All Years. Databases= SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH Lemmatization=On	Apr 2011: 84 Jul 2012: 12 Jan 2013: 2 Jan 2016: 9
5. Web of Science Core Collection (includes Conference Proceedings Citation Index; Thomson Reuters Web of Science) 1945 to 15 January 2016	Topic=(IQCODE OR "informant questionnaire on cognitive decline in the elderly" OR "IQ code") AND Topic=(dement* OR alzheimer* OR FTLD OR FTD OR "primary progressive aphasia" OR "progressive non-fluent aphasia" OR "frontotemporal lobar degeneration" OR "frontolobar degeneration" OR "frontal lobar degeneration" OR "pick* disease" OR "lewy bod*") Timespan=All Years. Databases= SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH Lemmatization=On	Apr 2011: 184 Jul 2012: 24 Jan 2013: 13 Jan 2016: 56
6. LILACS BIREME (Latin American and Caribbean Health Science Information database) (1982 to 15 January 2016)	"short-IQCODE" OR IQCODE OR "IQ code" OR "Informant Questionnaire" OR "Informant Questionnaires"	Apr 2011: 10 Jul 2012: 0 Jan 2013: 0 Jan 2016: 2
7. CINAHL EBSCO (Cumulative Index to Nursing and Allied Health Literature) (1982 to 15 January 2016)	S1 TX IQCODE S2 TX "informant questionnaire" S3 TX "IQ code" S4 TX screening instrument S5 S1 or S2 or S3 or S4 S6 (MM "Dementia+") S7 TX dement* S8 TX alzheimer* S9 S6 or S7 or S8 S10 S5 and S9	Apr 2011: 231 Jul 2012: 53 Jan 2013: 12 Jan 2016: 70

(*Continued*)

8. Additional review sources: • MEDION database (searched 15 January 2016 for all dates) • Database of Abstracts of Reviews of Effects (searched the Cochrane Library 2016, issue 1) • Health Technology Assessment Database (searched the Cochrane Library 2016, issue 1); • ARIF: Aggressive Research Intelligence Facility www.arif.bham.ac.uk (searched 31 January 2016 for all dates)		Jan 2013: 3 Jan 2016: 0
9 ALOIS (see Appendix 3 for the Medline strategy used to populate ALOIS) (searched 15 January 2016)		Jan 2013: 22 Jan 2016: 0
TOTAL before de-duplication of search results	Apr 2011: 1361 Jul 2012: 215 Jan 2013: 107 (+3 from additional review sources) Jan 2016: 149 TOTAL: 1835	
TOTAL after de-duplification and first-assess by the Trials Search Co-ordinator TOTAL after assessment of 220 by author team	220 83	

Appendix 3. Search strategy (MEDLINE OvidSP) run for specialised register (ALOIS)

MEDLINE in-process and other non-indexed citations and MEDLINE OvidSP (1950 to present)	1. “word recall”.ti,ab. 2. “7-minute screen”.ti,ab. 3. “6 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”.ti,ab. 22. “fuld 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. “neurobehavioural 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”.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).ti,ab.
81. “cognit* impair*”.ti,ab.
82. (cognit* adj4 (disorder* or declin* or fail* or function*)).ti,ab
83. (memory adj3 (complain* or declin* or function*)).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. 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. (animals not (humans and animals)).sh.
131. 129 not 130

The concepts for this are:

A Specific neuropsychological tests (lines 1-73)

B General terms (both free text and MeSH) for tests/testing/screening (lines 111-122)

C Outcome: dementia diagnosis (unfocused MeSH with diagnostic subheadings) (lines 107-109)

D Condition of interest: Dementia (general dementia terms both free text and MeSH - exploded and unfocused) (75-83)

E Methodological filter: not used to limit all search (85-105)

The concept combinations are:

1. (A OR B) AND C

	2. (A OR B) AND D AND E 3. A AND E **Search strategy (MEDLINE OvidSP) run for specialised register (ALOIS)** Search narrative: The search in Appendix 2 is largely based on a single concept: the index test (IQCODE). This is a sensitive approach to take. More complex and developed searches are run each month for the dementia group Every month the following strategy is run in MEDLINE (via OvidSP). The results are screened based on a reading of title and abstract. The full texts (where there is one) are then obtained and a few key details about each study are extracted including Index test/s and details of population and setting. For this review it was expected that most studies would be identified through a search of multiple sources based on one concept (the index test in question) . However, we felt it was worth also searching ALOIS for any studies which had evaluated the accuracy of IQCODE but had not referred to it in the title or abstract of the reference

Appendix 4. Assessment of reporting quality - STARDdem checklist

SECTION AND TOPIC		
TITLE/ABSTRACT KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity')
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups
METHODS		**METHODS**
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected. See also item 4 on recruitment and item 5 on sampling
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard? See also item 5 on sampling and item 16 on participant loss at each stage of the study
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected. See also item 4 on recruitment and item 16 on participant loss
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?

(*Continued*)

Test methods	7	The reference standard and its rationale.
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard. See also item 10 concerning the person(s) executing the tests
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard. See also item 8
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers. See also item 7
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals)
	13	Methods for calculating test reproducibility, if done.
RESULTS		**RESULTS**
Participants	14	When study was performed, including beginning and end dates of recruitment
	15	Clinical and demographic characteristics of the study population (at least information on age, sex, spectrum of presenting symptoms). See also item 18
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended). See also items 3 to 5
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard
	20	Any adverse events from performing the index tests or the reference standard
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals). See also item 12
	22	How indeterminate results, missing data and outliers of the index tests were handled

(*Continued*)

	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centres, if done
	24	Estimates of test reproducibility, if done. See also item 13
DISCUSSION		**DISCUSSION**
	25	Discuss the clinical applicability of the study findings.

Appendix 5. Assessment of methodological quality table QUADAS-2 tool

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 x 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 a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same 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?

(*Continued*)

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 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 6. Anchoring statements for quality assessment of IQCODE diagnostic studies

We provide some core anchoring statements for quality assessment of diagnostic test accuracy reviews of IQCODE in dementia. These statements are designed for use with the QUADAS-2 tool and were derived during a two-day, multidisciplinary focus group.

During the focus group and the piloting and validation of this guidance, it was clear that certain issues were key to assessing quality, while other issues were important to record, but less important for assessing overall quality. To assist, we describe a system wherein certain items can dominate. For these dominant items, if scored 'high risk', then that section of the QUADAS-2 results table is likely to be scored as high risk of bias, regardless of other scores. For example, in dementia diagnostic test accuracy studies, ensuring that clinicians performing the dementia assessment are blinded to the results of index test is fundamental. If this blinding was not present, then the item on the reference standard should be scored 'high risk' of bias, regardless of the other contributory elements.

We have detailed how QUADAS2 has been operationlised for use with dementia reference standard studies below. In these descriptors, dominant items are labelled as 'hIgh risk'.

In assessing individual items, the score of 'unclear' should only be given 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

Patient selection

1. Was a case-control or similar design avoided?

Designs similar to case control that may introduce bias are those designs where the study team deliberately increase or decrease the proportion of patients with the target condition. For example, a population study may be enriched with extra dementia patients from a secondary care setting. Such studies will be automatically labelled high risk of bias and will be assessed as a potential source of heterogeneity.

High risk of bias (in fact, case-control studies will not be included in this review)

2. Was the sampling method appropriate?

Where sampling is used, the designs least likely to cause bias are consecutive sampling or random sampling. Sampling that is based on volunteers or selecting participants from a clinic or research resource is prone to bias.

High risk of bias

3. Are exclusion criteria described and appropriate?

The study will be automatically graded as unclear if exclusions are not detailed (pending contact with study authors). Where exclusions are detailed, the study will be graded as low risk if the review authors feel the exclusions are appropriate. Certain exclusions common to many studies of dementia are: medical instability, terminal disease, alcohol or substance misuse, concomitant psychiatric diagnosis, or other neurodegenerative conditions. For a community sample, we would expect relatively few exclusions.

Post hoc exclusions will be labelled 'high risk' of bias.
Low risk

Index test

4. Was IQCODE assessment performed without knowledge of clinical dementia diagnosis?

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 explicitly described, or if there is a clear temporal pattern to the order of testing that precludes the need for formal blinding i.e. all IQCODE assessments were performed before dementia assessment.
High risk

5. Were IQCODE thresholds pre-specified?

For scales, there is often a reference point (in units or categories) above which participants are classified as 'test positive'; this may be referred to as threshold; clinical cut-off or dichotomisation point. A study is classified to be at high 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.
Low risk

6. Were sufficient data on IQCODE application given for the test to be repeated in an independent study?

Particular points of interest for IQCODE include method of administration (for example, self-completed questionnaire versus direct questioning interview), nature of informant, and language of assessment. If a novel form of IQCODE is used, details of the scale should be included, or a reference given to an appropriate descriptive text. Where IQCODE is used in a novel manner, for example, a translated questionnaire, there should be evidence of validation.
Low risk

Reference standard

7. Is the assessment used for clinical diagnosis of dementia acceptable?

Commonly used international criteria to 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 dementias, and the NINDS-AIREN criteria for vascular dementia. Where 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.
High risk

8. Was clinical assessment for dementia performed without knowledge of IQCODE?

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 there is a clear temporal pattern to the order of testing i.e. all dementia assessments are performed before IQCODE 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. We have pre-specified that a dementia diagnosis that explicitly uses IQCDODE will be classified as high risk of bias.
High risk

9. Were sufficient data on dementia assessment method given for the assessment to be repeated in an independent study?

The criteria used for clinical assessment are discussed in another item. Particular points of interest for dementia assessment include the background of the assessor, training/expertise of the assessor, and additional information available to inform the diagnosis (neuroimaging; neuropsychological testing).
Low risk

Patient flow

10. Was there an appropriate interval between IQCODE and clinical dementia assessment.

For a study looking at delayed verification, there is no agreement on how long the interval should be between index test and first/last assessment for dementia. An interval of less than six months is unlikely to be sufficient time for progression.
Low risk of bias

11. Did all patients get the same assessment for dementia regardless of IQCODE result?

There may be scenarios where only those patients who score 'test positive' on IQCODE have a more detailed assessment. Where dementia assessment (or other reference standard) differs between patients, this should be classified as high risk of bias.
High risk of bias

12. Were all patients who received IQCODE assessment included in the final analysis?

If dropouts, these should be accounted for; a maximum proportion of dropouts for this domain to remain at low risk of bias has been specified as 20%.
Low risk of bias

13. Were missing IQCODE results or un-interpretable IQCODE results reported?

Where missing results are reported, and if there is substantial attrition (we have set an arbitrary value of 50% missing data), this should be scored as high risk of bias.
Low risk of bias

Applicability

14. Were included patients representative of the general population of interest?

The included patients should match the intended population, as described in the review question. If not already specified in the review inclusion criteria, setting will be particularly important - the review authors should consider the population in terms of symptoms, pre-testing, and potential disease prevalence. Studies that use very selected patients or subgroups will be classified as poor applicability.

15. Was IQCODE performed consistently and in a manner similar to its use in clinical practice?

IQCODE studies will be judged against the original description of its use.

16. Was clinical diagnosis of dementia (or other reference standard) made in a manner similar to current clinical practice?

For many reviews, inclusion criteria and assessment for risk of bias will already have assessed the dementia diagnosis. For certain reviews, an applicability statement relating to the reference standard may not be applicable. There is the possibility that a form of dementia assessment, although valid, may diagnose a far larger proportion of patients with disease than usual clinical practice. In this instance, the item should be rated as poor applicability.

Appendix 7. STARDdem (reporting quality) results

Study ID	STARDdem Item Assessment	
	Yes	No
Caratozzolo 2014	3,4,6,9,14,15,16,17	1,2,5,7,8,10,11,12,13,18,19,20,21,22,23,24,25
Henon 2001	3,4,5,6,7,8,9,14,15,17	1,2,10,11,12,13,16,18,19,20,21,22,23,24,25
Krogseth 2011	3,4,6,7,8,9,10,14,15,16,17	1,2,5,11,12,13,18,19,20,21,22,23,24,25

CONTRIBUTIONS OF AUTHORS

JKH drafted the initial manuscript and assisted with data extraction, quality assessment, and analysis.

DJS and RM provided supervision and input to the protocol and review.

ANS assisted with the search strategy and searching, and provided input to the protocol and review.

RSS-P assisted with data extraction, quality assessment, and analysis.

TJQ drafted the protocol, and assisted with searching, data extraction, quality assessment, and analysis.

DECLARATIONS OF INTEREST

No relevant disclosures or conflicts of interest for the content of this review.

SOURCES OF SUPPORT

Internal sources

- The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology & The Alzheimer Scotland Dementia Research Centre, UK.

JKH is supported by a Clinical Research Fellowship funded by Alzheimer Scotland and The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/L501530/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged.

- The Stroke Association & The Chief Scientist Office for Scotland, UK.

TJQ is supported by a joint Stroke Association/Chief Scientist Office Senior Clinical Lectureship.

External sources

- NIHR, UK.

This review was supported by the National Institute for Health Research, via a Cochrane Programme Grant to the Cochrane Dementia and Cognitive Improvement group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health, UK.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Quantitative analysis and planned sensitivity analyses were not possible due to the heterogeneity of the included studies.

INDEX TERMS

Medical Subject Headings (MeSH)

*Early Diagnosis; *Surveys and Questionnaires; Cognition Disorders [*diagnosis]; Cohort Studies; Dementia [*diagnosis]; Hip Fractures; Sensitivity and Specificity; Stroke; Time Factors

MeSH check words

Aged; Humans