Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations (Review)

Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ

THE COCHRANE COLLABORATION®

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2014, Issue 4

http://www.thecochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations

Terry J Quinn[1], Patricia Fearon[2], Anna H Noel-Storr[3], Camilla Young[2], Rupert McShane[3], David J Stott[2]

[1]Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [2]Academic Section of Geriatric Medicine, University of Glasgow, Glasgow, UK. [3]Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Contact address: Terry J Quinn, Cardiovascular and Medical Sciences, University of Glasgow, Walton Building, Glasgow Royal Infirmary, Glasgow, G4 0SF, UK. Terry.Quinn@glasgow.ac.uk.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.
Publication status and date: New, published in Issue 4, 2014.
Review content assessed as up-to-date: 1 February 2013.

Citation: 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. Art. No.: CD010079. DOI: 10.1002/14651858.CD010079.pub2.

ABSTRACT

Background

Various tools exist for initial assessment of possible dementia with no consensus on the optimal assessment method. Instruments that use collateral sources to assess change in cognitive function over time may have particular utility. The most commonly used informant dementia assessment is the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE).

A synthesis of the available data regarding IQCODE accuracy will help inform cognitive assessment strategies for clinical practice, research and policy.

Objectives

Our primary objective was to determine the diagnostic accuracy of the informant based questionnaire IQCODE, for detection of all cause (undifferentiated) dementia in community-dwelling adults with no previous cognitive assessment. We sought to describe the accuracy of IQCODE (the index test) against a clinical diagnosis of dementia (the reference standard).

Our secondary objective was to describe the effect of heterogeneity on the summary estimates. We were particularly interested in the traditional 26-item scale versus the 16-item short form; and language of administration. We explored the effect of varying the threshold IQCODE score used to define 'test positivity'.

Search methods

We searched the following sources on 28 January 2013: ALOIS (Cochrane Dementia and Cognitive Improvement Group), MEDLINE (OvidSP), EMBASE (OvidSP), PsycINFO (OvidSP), BIOSIS Previews (ISI Web of Knowledge), Web of Science with Conference Proceedings (ISI Web of Knowledge), LILACS (BIREME). We also searched sources relevant or specific to diagnostic test accuracy: MEDION (Universities of Maastrict and Leuven); DARE (York University); ARIF (Birmingham University). We used sensitive search terms based on MeSH terms and other controlled vocabulary.

Selection criteria

We selected those studies performed in community settings that used (not necessarily exclusively) the IQCODE to assess for presence of dementia and, where dementia diagnosis was confirmed, with clinical assessment. Our intention with limiting the search to a 'community' setting was to include those studies closest to population level assessment. Within our predefined community inclusion criteria, there were relevant papers that fulfilled our definition of community dwelling but represented a selected population, for example stroke survivors. We included these studies but performed sensitivity analyses to assess the effects of these less representative populations on the summary results.

Data collection and analysis

We screened all titles generated by the electronic database searches and abstracts of all potentially relevant studies were reviewed. Full papers were assessed for eligibility and data extracted by two independent assessors. For quality assessment (risk of bias and applicability) we used the QUADAS 2 tool. We included test accuracy data on the IQCODE used at predefined diagnostic thresholds. Where data allowed, we performed meta-analyses to calculate summary values of sensitivity and specificity with corresponding 95% confidence intervals (CIs). We pre-specified analyses to describe the effect of IQCODE format (traditional or short form) and language of administration for the IQCODE.

Main results

From 16,144 citations, 71 papers described IQCODE test accuracy. We included 10 papers (11 independent datasets) representing data from 2644 individuals (n = 379 (14%) with dementia). Using IQCODE cut-offs commonly employed in clinical practice (3.3, 3.4, 3.5, 3.6) the sensitivity and specificity of IQCODE for diagnosis of dementia across the studies were generally above 75%.

Taking an IQCODE threshold of 3.3 (or closest available) the sensitivity was 0.80 (95% CI 0.75 to 0.85); specificity was 0.84 (95% CI 0.78 to 0.90); positive likelihood ratio was 5.2 (95% CI 3.7 to 7.5) and the negative likelihood ratio was 0.23 (95% CI 0.19 to 0.29).

Comparative analysis suggested no significant difference in the test accuracy of the 16 and 26-item IQCODE tests and no significant difference in test accuracy by language of administration. There was little difference in sensitivity across our predefined diagnostic cut-points.

There was substantial heterogeneity in the included studies. Sensitivity analyses removing potentially unrepresentative populations in these studies made little difference to the pooled data estimates.

The majority of included papers had potential for bias, particularly around participant selection and sampling. The quality of reporting was suboptimal particularly regarding timing of assessments and descriptors of reproducibility and inter-observer variability.

Authors' conclusions

Published data suggest that if using the IQCODE for community dwelling older adults, the 16 item IQCODE may be preferable to the traditional scale due to lesser test burden and no obvious difference in accuracy. Although IQCODE test accuracy is in a range that many would consider 'reasonable', in the context of community or population settings the use of the IQCODE alone would result in substantial misdiagnosis and false reassurance. Across the included studies there were issues with heterogeneity, several potential biases and suboptimal reporting quality.

BACKGROUND

Dementia is a substantial and growing public health concern (Ferri 2005). Depending on the case definition employed, contemporary estimates of dementia prevalence in the United States are in the range of 2.5 to 4.5 million individuals (Hebert 2003). Dementia is predominantly a disease of older adults, with a 5% prevalence in adults aged over 60 years, increasing to up to 50% in adults aged over 85 years (Ferri 2005). Changes in population demographics will result in increased absolute and proportional numbers of older adults and will be accompanied by increases in dementia incidence and prevalence, albeit the extent of this increase is debated (Matthews 2013). Dementia is not limited to 'Western' nations and an increasing prevalence is particularly marked in countries such as China and India (Ferri 2005).

Given the projected global increase in dementia prevalence, there is a potential tension between the clinical requirements for robust diagnosis at the individual patient level and the need for equitable, easy access to diagnosis at a population level. The ideal would be expert, multidisciplinary assessment informed by various supplementary investigations. Such an approach may be possible in a secondary or tertiary care setting, however, in a community or primary care setting the population is too large and the prevalence of the disease will be low relative to the more specialist memory-clinic setting.

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

Initial assessment often takes the form of brief, direct cognitive testing. Using this method a single test can 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 over a period of years and so an attractive approach is to question collateral sources with sufficient knowledge of the patient. Various terms have been used to describe the person(s) providing descriptions of the patient's cognition including proxy, collateral, informant, carer etc. We should make no assumptions about the relationship of the person providing the description and for consistency throughout the text we use the term informant.

Informant-based interviews have been described that aim to retrospectively assess change in function over a period of time. An instrument prevalent in research and clinical practice is the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE), and this is the focus of our review.

There is no consensus on the optimal test for dementia and choice of test 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 short dementia 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.

Target condition being diagnosed

The target condition for this diagnostic test accuracy review is dementia (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 and at present there is no 'cure', although numerous interventions to slow or arrest cognitive decline have been studied.(Birks 2006; Clare 2003; McShane 2006).

Dementia remains a clinical diagnosis, based on a history from the patient and suitable informant sources and direct examination including cognitive assessment. Expert committees have described criteria for diagnosis of the dementia syndrome and its various subtypes (Erkinjuntti 2000; McKeith 2005; McKhann 1984; McKhann 2001; Roman 1993). Various clinical diagnostic protocols are available and although there are slight variations in European and American guidance, core features are common to all diagnostic criteria (McKhann 2011) (Appendix 1).

We recognise that there is no universally accepted, gold standard dementia diagnostic strategy. We chose expert clinical diagnosis as our gold standard (reference standard) for describing IQCODE accuracy as we believe this is most in keeping with current diagnostic criteria and best practice. Previous studies have used neuropathology as a gold standard. For the purpose of testing diagnostic accuracy in large unselected populations, limiting analysis to those studies with neuropathological confirmed diagnosis is likely to yield limited and highly selected data (Savva 2009). Criteria for diagnosis of dementia are evolving in line with improvements in our understanding of the underlying pathophysiological processes. Various biomarkers based on biological fluid assays or functional and quantitative neuroimaging have shown promise but to date they are not accepted or validated as independent diagnostic tests (McKhann 2011; Noel-Storr 2012).

The label of dementia encompasses varying pathologies of which Alzheimer's disease is the most common. For our reference standard of clinical diagnosis, we accept a dementia diagnosis made according to any of the internationally accepted diagnostic criteria, with exemplars being the various iterations of 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 (Appendix 1). We also recognise the various diagnostic criteria 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); McKeith criteria for Lewy Body dementia (McKeith 2005); Lund criteria for frontotemporal dementias (McKhann 2001); and the National Institute of Neurological Disorders and Stroke - Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria for vascular dementia (Roman 1993). Diagnostic criteria are continually evolving in line with a better clinical and scientific under-

standing of dementia, for example at the time of review the fifth edition of DSM was in pre-release.

Index test(s)

We chose the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (Jorm 1988) as our index test of interest. The IQCODE was originally described as a 26-item informant questionnaire designed to retrospectively ascertain change in cognitive and functional performance over a 10-year time period (Jorm 1988). IQCODE is designed as a brief assessment 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 IQCODE is prevalent in both clinical practice and research (Holsinger 2007) and the questionnaire has been translated into several languages (www.anu.edu.au/iqcode/). IQCODE has a number of features that make it attractive for clinical and research use. The questions used have an immediacy and relevance that is likely to appeal to users. Assessment and (informant) scoring takes around five to seven minutes and as the scale is not typically interviewer administered it requires minimal training in application and scoring (Holsinger 2007). Proponents of IQCODE suggest several potentially favourable properties of the IQCODE when compared to standard direct assessments. The IQCODE may be less prone to bias from cultural norms and previous levels of education; the scale has good inter-rater reliability; and internal consistency is uniformly high with Cronbach's alpha in the range 0.93 to 0.97 (Jorm 1989A). Validation work has included validation against measures of cognitive change; neuropathology; neuroimaging and neuropsychological assessment (Cordoliani-Mackowiak 2003; Jorm 2000A; Rockwood 1998).

A shortened 16-item version is available; this modified IQCODE is common in clinical practice and has been recommended as the preferred IQCODE format (Jorm 2004). Further modifications to IQCODE are described including fewer items and assessment over shorter time periods. For our analysis we chose to include all versions of IQCODE but present results for the original and modified scales separately in the first instance. In this text, the term IQCODE refers to the original 26-item questionnaire as described by Jorm (Jorm 1988).

IQCODE cut-off scores used to define test positivity vary with the demographics of the population and the reason for testing. In the original development and validation work, normative data were described with a total score of > 93 or average score of > 3.31 indicative of cognitive impairment (Jorm 1988). There is no consensus on the optimal threshold and various authors have described improved diagnostic accuracy with other cut-offs.

The full 26 and 16-item versions of ICQODE with scoring rules are available as appendices (Appendix 2; Appendix 3).

Clinical pathway

A key element of effective management in dementia is robust diagnosis. Recent guidelines place emphasis on early diagnosis to facilitate improved management and to allow informed discussions and planning with patients and carers. The utility of screening for an early, unprompted diagnosis of dementia remains a subject of debate. There are major pressures for early diagnosis from third sector organisations, patient representative groups, and the pharmaceutical industry; and in certain countries opportunistic cognitive screening or case-finding is suggested (Brunet 2012; Cordell 2013).

We recognise the importance of healthcare setting and populations in describing test properties. We have defined a series of settings and populations for reviews; these are based on the reason for performing the index (IQCODE) test and the likely prevalence of dementia.

Studies can be based in secondary care, that is where a referral has already been made by a healthcare professional and where there may have been some form of cognitive screening or selection.

In the general practice or primary care setting, patients generally self-present to a non-specialist service because of subjective memory complaints, usually with no prior cognitive testing. In this setting the purpose of cognitive testing is to triag' individuals to inform decisions about onward specialist referral.

A study in a community (population) setting will generally be an unselected cohort with no previous cognitive assessment. The purpose of community cognitive testing may be population screening, or to inform epidemiological studies. Our intention with the community setting was to include those studies closest to population level screening. Methodologies for selecting representative community samples differ and for this review we adopted an inclusive approach in the first instance, including studies where the populations were community dwelling and not selected on the basis of cognitive scores or symptoms. We would expect lower prevalence of disease in the community setting compared to other settings. This is an important methodological point as in certain studies researchers 'enrich' a community population with dementia cases. This process can artificially improve test accuracy and does not allow for description of those metrics that relate to population prevalence, for example positive or negative predictive value.

For this review we described the test accuracy of IQCODE when used in a community setting. For consistency through the review we have used the term community. Reviews describing studies in other settings will also be available in due course.

Prior test(s)

For a review in a community setting, we would expect that the majority of individuals included will have had no previous assessment for cognitive problems. We did not include studies where recruitment was based on results of previous cognitive test(s).

Role of index test(s)

Although we use the term diagnosis in this review, we recognise that in practice IQCODE alone is not sufficient to make a diagnosis. Rather IQCODE is used as an initial case finding, triage, or screening test that can inform the need for further assessment or assist with diagnosis in conjunction with direct patient assessment and investigations. For ease of understanding and consistency with other reviews we used the term diagnostic accuracy to infer 'accuracy of IQCODE test for suggesting a possible dementia case'.

Alternative test(s)

Several other dementia assessment tools have been described, these are usually performance-based measures that rely on comparing single or multi-domain cognitive testing against population-specific normative data (Brodaty 2002; Burns 2004; Holsinger 2007). There are fewer informant interviews available. An alternative to IQCODE that is popular in North America is the eight-item Interview to Differentiate Aging and Dementia (AD-8) (Galvin 2005).

For this review we did not consider other cognitive screening or assessment tools and have chosen not to include other tests as comparators. Currently there is no standard practice biomarker or neuropsychological test and so we felt that making decisions on meaningful comparators was premature. Where a paper describes IQCODE with in-study comparison against another tool, we included the IQCODE data only. Where the IQCODE code was used in combination with another tool, we included the IQCODE data only.

Our IQCODE diagnostic studies form part of a larger body of work describing the test accuracy of all commonly used scales and Cochrane diagnostic test accuracy (DTA) reviews specific to the AD-8; Abbreviated Mental Test (AMT); Clock Drawing Test (CDT); Mini-Cognitive Assessment Instrument (Mini-Cog); Mini-Mental State Examination (MMSE); Montreal Cognitive Assessment (MoCA) and General Practitioner assessment of Cognition (GPcog) are planned or in production (Appendix 4).

Rationale

Clinical properties of a dementia test should not be assumed and formal testing of sensitivity, specificity and other properties of IQCODE should be performed and collated before the tool can be recommended.

IQCODE is commonly used in practice and research; it is used internationally and is one of only a few validated informant-based tools. Literature describing test accuracy of IQCODE in different settings is available, although some of these studies have been modest in size. Thus a systematic review and, if possible, meta-analysis of the diagnostic test accuracy of IQCODE is warranted.

OBJECTIVES

Our primary objective was to determine the diagnostic accuracy of the informant based questionnaire IQCODE, for detection of all cause (undifferentiated) dementia in community-dwelling adults with no previous cognitive assessment. We sought to describe the accuracy of IQCODE (the index test) against a clinical diagnosis of dementia (the reference standard).

Secondary objectives

Where data were available, we planned to describe the following.

1. The diagnostic accuracy of IQCODE at various pre-specified thresholds. We recognize that various thresholds or cut-off scores have been used to define IQCODE test positive states. We described the test accuracy of IQCODE for the following cut-off scores (rounded where necessary): 3.3, 3.4, 3.5, 3.6. These thresholds have been chosen to represent the range of cut-offs that are commonly used in practice and research; we have been inclusive in our choice of cut-off to maximize available data for review.
2. Accuracy of IQCODE for diagnosis of the commonest specific dementia subtype, Alzheimer's disease dementia.
3. Effects of heterogeneity (see below) on the reported diagnostic accuracy of IQCODE.

Our focused study question, restricting this review to a community setting, was designed to remove potential heterogeneity relating to study design and setting. Other sources of heterogeneity in dementia studies such as treatment, intervention or duration of follow-up are not applicable to this review and were considered within the inclusion and exclusion criteria. The properties of a tool describe the behaviour of the instrument under particular circumstances. Thus for our assessment of potential sources of heterogeneity (where data allowed), we collated data on key features of the study population namely age; features of the index test, namely language of administration and IQCODE format; features of the reference standard, namely diagnostic criteria used; and diagnostic methodology.

METHODS

Criteria for considering studies for this review

Types of studies

All studies of community-based cohorts were potentially eligible for the review. As discussed, we used the rubric 'community' to include studies of community dwelling older adults, unselected on the basis of cognitive scores or symptoms.

Many studies that assess test properties use a case control methodology. This approach is prone to a number of potential biases and may give artificially high values for test accuracy. For certain studies, in particular where populations are 'enriched' with dementia cases, case control methodology may be employed but not explicitly stated. We elected to include potential case control studies in our initial screening review of the search results and then assess studies on a case by case basis. Where case control or study enriching was employed we did not include these in the summary data or pool these data with other studies.

Case studies or samples with very small numbers (chosen as 10 participants or less for the purposes of this review) were not included.

Participants

All community-dwelling adults (aged over 18 years) were potentially eligible. We suspected that the majority of included participants in the eligible studies would be aged over 65 years.

Our definition of a community-based study setting was a study where participants were community dwelling, had not been referred, had not had extensive cognitive testing and had not self-presented for assessment of subjective memory problems. We anticipated that studies would largely be of unselected community-dwelling adults; this cohort is itself heterogeneous. We did not predefine exclusion criteria relating to the 'case-mix' of the population studied but assessed applicability for each study. Where a population was community dwelling and unselected on the basis of cognition but was potentially not representative of the population, for example a study with a focus on stroke-survivors, we chose to explore the effect of these studies on the findings using sensitivity analyses.

Index tests

Studies had to include (not necessarily exclusively) IQCODE used as an informant questionnaire.

IQCODE has been translated into various languages. The properties of a translated IQCODE in a cohort of non-English speakers may differ from the properties of the original English language questionnaire. We collected data on the principle language used for IQCODE assessment in studies to allow for assessment of heterogeneity in relation to language.

Since its original description modifications to the administration of IQCODE have been described (Jorm 2004). Shorter forms of informant questionnaires that test fewer domains are available and properties may differ from the original 26-item IQCODE tool. We included all such versions of IQCODE but presented separate analysis limited to the commonest 26 and 16-item versions. A modified IQCODE for self-assessment has been described. As our interest was informant interviews, self-assessment IQCODE was not included in the review (Cullen 2007).

Target conditions

Papers reporting any clinical diagnosis of all cause (unspecified) dementia were potentially eligible for inclusion. Defining a particular dementia subtype was not required although where available, these data were recorded.

Reference standards

Our reference standard was clinical diagnosis of 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. Our definition of clinical diagnosis included all cause (unspecified) dementia, using any recognised diagnostic criteria (for example ICD-10; DSM-IV). The dementia diagnosis could specify a pathological subtype and all dementia subtypes were included (examples McKeith 2005; McKhann 1984; McKhann 2001; Roman 1993). Clinicians may have used imaging, pathology or other data to aid diagnosis, however, diagnosis based only on these data without corresponding clinical assessment were not included. We recognise that different iterations of diagnostic criteria may not be directly comparable and that diagnosis may vary with the degree or manner in which the criteria have been operationalised (for example individual clinician versus algorithm versus consensus determination) and so data on method and application of dementia diagnosis was collected for each study.

We did not set criteria relating to severity or stage of dementia diagnosis, instead any clinical diagnosis of dementia (not mild cognitive impairment or its equivalents) was classified. We planned to explore stage or severity of dementia as a potential source of heterogeneity.

Search methods for identification of studies

We used a variety of information sources to ensure all relevant studies were included. Terms for electronic database searching were devised in conjunction with the Trials Search Co-ordinator 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 dementia test accuracy studies. The output of the searches was then assessed to

select those papers that could be pertinent to IQCODE, with further selection for directly relevant papers and those papers with a community (population) focus.

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 (OvidSP), ISI Web of Science and Conference Proceedings (ISI Web of Knowledge), CINAHL (EBSCOhost) and LILACS (BIREME). See Appendix 5 and Appendix 6 for the search strategies. The final search date was 28 January 2013.

We also searched sources specific to diagnostic accuracy and healthcare assessment:

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

We did not apply any language or date restrictions to the electronic searches. Translation services were used as necessary.

Initial screening of the search results was performed by a single researcher from the Cochrane Dementia and Cognitive Impairment Group with extensive experience of systematic reviews (ANS). All subsequent assessments of search results, based either on assessment of titles, abstracts or full text, were performed by independent paired assessors (TQ, PF).

Searching other resources

Grey literature: 'grey' literature was identified through searching conference proceedings on EMBASE (OvidSP) and through the ISI Web of Knowledge platform.

Handsearching: we did not perform handsearching. The evidence base on handsearching for DTAs is not yet known and there is no clear guidance on whether handsearching is worthwhile.

Reference lists: we checked the reference lists of all relevant studies and reviews in the field for further possible titles and repeated the process until no new titles were found (Greenhalgh 2005).

Correspondence: we contacted research groups who have published or are conducting work on the IQCODE for dementia diagnosis, informed by results of the initial search.

Relevant additional studies were searched for in PubMed using the related article feature. Relevant studies were examined in the citation databases of Science Citation Index and Scopus to ascertain any further relevant studies.

Data collection and analysis

Selection of studies

One review author (ANS) screened for relevance all titles generated by the initial electronic database searches. The initial search was a sensitive, generic search designed to include many potential dementia screening tools. Titles potentially relevant to IQCODE were selected by two review authors (ANS, TQ). All further reviews of studies and selection were performed by two independent researchers (TQ, PF). The potential IQCODE related titles were reviewed and all eligible studies were assessed as abstracts; potentially relevant studies were assessed against inclusion criteria as full manuscripts. Disagreement was resolved by discussion, with potential to involve a third author (DJS) as arbitrator if necessary.

We adopted a hierarchical approach to exclusion, first excluding studies on the basis of index test and reference standard and then on the basis of sample size and study data. Finally, we assessed all IQCODE papers with regard to setting.

Where a study may have included useable data but these were not presented in the published manuscript (labelled as data not suitable for analysis on flowchart), we contacted the authors directly to request further information. 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

Data were extracted to a study-specific pro forma that included clinical and demographic details of the participants; details of IQCODE administration; and details of the dementia diagnosis process. The pro forma was piloted against two of the included papers before use.

Where IQCODE data were given for a number of cut-points, we extracted data for each of our pre-specified cut-points: 3.3, 3.4, 3.5, 3.6. Where thresholds were described to two decimal places, we chose the cut-point closest to the point of interest (that is all scores less than 3.35 would be scored as 3.3; all scores 3.35 or greater would be scored as 3.4. Data were extracted to a standard two by two table.

Data extraction was performed independently by review authors (TQ, PF). Authors were based in differing centres and were blinded to each other's data until extraction was complete. Data pro formas were then compared and discussed with reference to the original papers. Disagreement in data extraction was resolved by discussion, with the potential to involve a third author (DJS) as arbitrator if necessary.

For each included paper, the flow of participants (numbers recruited, included, assessed) was detailed 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.

Quality of study reporting was assessed using the STARD checklist (Bossuyt 2003) (Appendix 7). We recognise that a dementia-specific extension to complement STARD (STARDdem) (http://starddem.org/) is proposed, however the content of STARDdem was not finalised at the time of this analysis. STARD data were tabulated and presented as an appendix to the review.

We assessed the methodological quality of each study using the QUADAS-2 tool (http://www.bris.ac.uk/quadas/quadas-2) (Appendix 8). This tool incorporates domains specific to patient selection; index test; reference standard; and patient 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 8. 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 (Appendix 9).

Both assessments were performed by paired independent raters (TJQ, PF) who were blinded to each other's scores. Disagreement was resolved by further review and discussion with the potential to involve a third author (DJS) as arbitrator if necessary.

QUADAS-2 data were not used to form a summary quality score, rather we chose to present a narrative summary describing the numbers of studies that found high, low or unclear risk of bias and 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 dichotomous variable dementia or no dementia. Thus, we applied the current DTA framework for analysis of a single test to fit the extracted data to a standard two by two table showing binary test results cross-classified with the binary reference standard. This process was repeated for each of our pre-specified IQCODE threshold scores.

We used RevMan 5.2 (RevMan 2011) to calculate sensitivity, specificity and their 95% confidence intervals (CIs) from the two by two tables abstracted from the included studies. These data were presented graphically in forest plots to allow basic visual inspection of individual studies only. Standard forest plots with graphical representation of summary estimates are not suited to quantitative synthesis of DTA data. Using software additional to RevMan (SAS release 9.1) we used the bivariate method to calculate summary values within each pre-specified cut-off. The bivariate methods (Reitsma 2005) enabled us to calculate summary estimates of sensitivity and specificity while correctly dealing with the different sources of variation: (1) imprecision by which sensitivity and specificity have been measured within each study; (2) variation beyond chance in sensitivity and specificity between studies; (3) any correlation that might exist between sensitivity and specificity. The results for each chosen threshold were described as sensitivity and specificity and all accuracy measures were estimated with their 95% CI. Where data allowed, we chose to present individual study results graphically by plotting estimates of sensitivities and specificities in the receiver operating characteristic (ROC) space. We also described metrics of pooled positive and negative likelihood ratios. To allow an overview of IQCODE test accuracy we performed a further analysis: pooling data at a common threshold (3.3 or closest), chosen to maximise the data available for inclusion.

The presence of statistical heterogeneity was assessed by visual inspection of the included study results plotted in the ROC space relative to the putative summary accuracy estimates.

Investigations of heterogeneity

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

The properties of a tool describe the behaviour of the instrument under particular circumstances. Thus, for our assessment of potential sources of heterogeneity (where data allowed) we collected data to inform two broad pre-specified areas of interest. These were:

- clinical criteria used to reach dementia diagnosis (for example ICD-10; DSM-IV) and the methodology used to reach the dementia diagnosis (e.g. individual assessment; group (consensus) assessment);
- technical features of the testing strategy (version of IQCODE (language); numbers of items, that is short form of IQCODE or long form).

Where data allowed we performed pooled analysis with these factors as covariates and compared results of subgroups. We pre-specified that we would present data from the traditional (26 questions) and short form (16 questions) IQCODE separately.

Sensitivity analyses

Where appropriate (that is if not already explored in our analyses of heterogeneity) and as data allowed, we planned to explore the sensitivity of any summary accuracy estimates to aspects of study quality guided by the anchoring statements developed in our QUADAS-2 exercise. We pre-specified sensitivity analysis where we planned to exclude studies of low quality (high likelihood of bias) to determine if the results are influenced by inclusion of the lower quality studies; and sensitivity analysis excluding studies that may have unrepresentative populations.

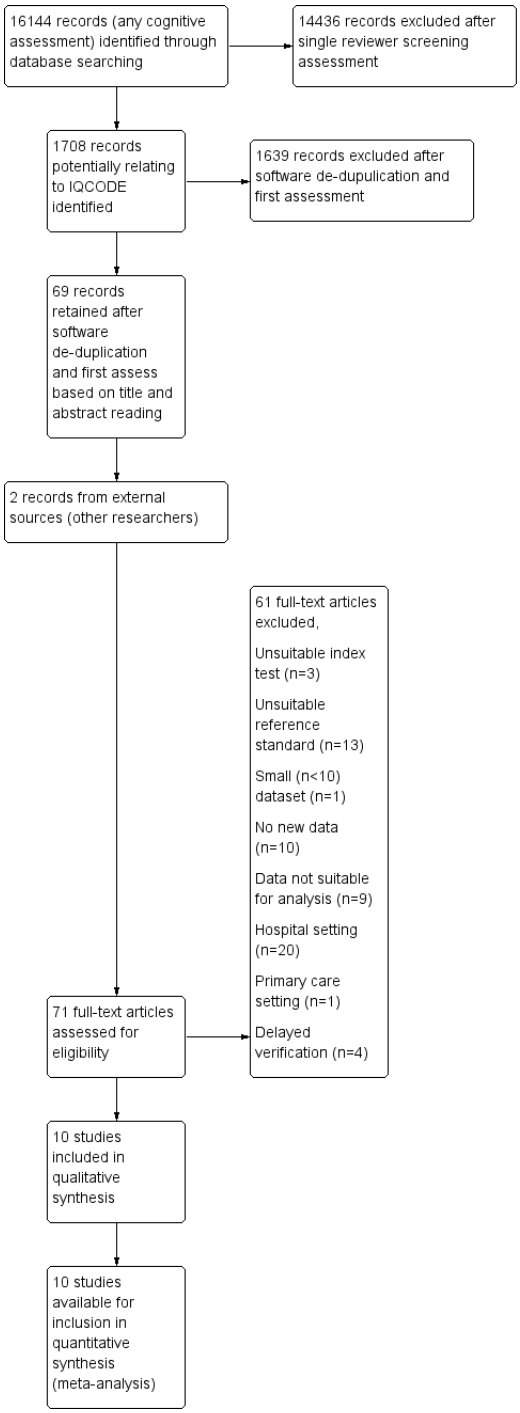
RESULTS

Results of the search

Our search resulted in 16144 citations, of which 71 full text papers were assessed for eligibility.

We excluded 61 papers (Figure 1). Reasons for exclusion were: population not from a community (population) setting; no IQCODE data or unsuitable IQCODE data; small numbers of included participants; no clinical diagnosis of dementia; repeat data sets; data not suitable for analysis (Characteristics of excluded studies).

Figure 1. Study flow diagram.



Eight studies that were identified required translation, these papers were not suitable for this review but the data have been used for reviews of IQCODE in other healthcare settings. We contacted 14 authors to provide useable data, of whom 10 responded. These data were not suitable for this (community setting) review but have been used for other IQCODE analyses in this family of reviews (see Acknowledgements).

This review included 10 studies representing 11 data sets (n = 2644 participants) (Summary of findings 1; Summary of findings 2; Summary of findings 3).

Methodological quality of included studies

We described risk of bias using the QUADAS 2 methodology (Appendix 8).

No study was graded low risk of bias for all the categories of QUADAS-2 (Figure 2; Figure 3). Areas of particular concern for bias were around: participant sampling procedures (n = 2 papers graded low risk, with few papers using a true consecutive sampling frame) and application of index test (n = 1 paper graded low risk of bias, with most papers giving insufficient detail on how the IQCODE was actually applied in practice). There were also concerns around applicability, particularly concerning patient selection procedures (n = 1 paper graded no concern, with few studies recruiting a cohort representative of community-dwelling older adults).

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

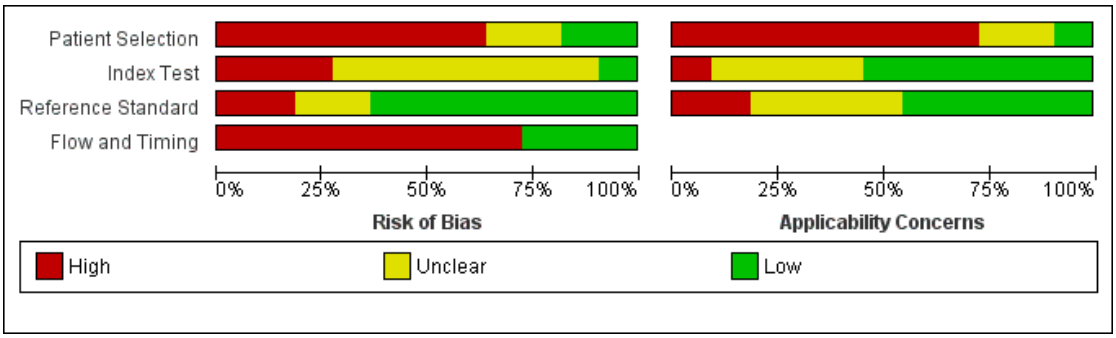
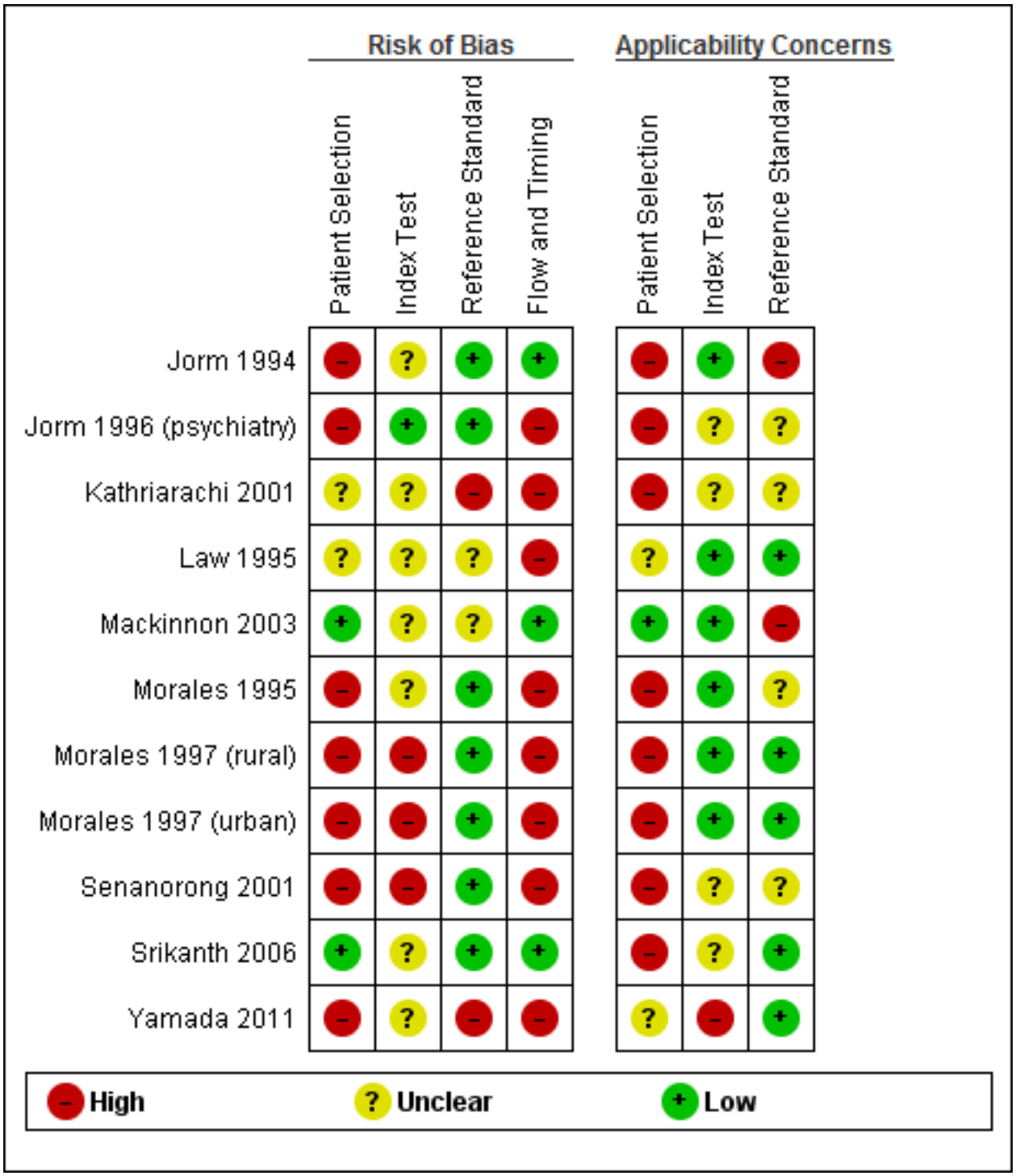


Figure 3. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.



We described reporting quality using the STARD guidance (Appendix 7). One paper (Yamada 2011) was not included in this process as we had an expanded conference abstract or poster and a paper describing the study methodology (Yamada 2008) but a full manuscript with IQCODE data had not yet been published. There were limitations in reporting across all included papers (Appendix 10). No paper included all the details recommended in the STARD statement; particular areas of study reporting that could be improved were: reporting of timing of the index test and reference standard (n = 1 paper reported when the IQCODE was performed in relation to the diagnostic evaluation); handling of indeterminate results (n = 0 papers reported, for example, how incomplete IQCODE questionnaires were handled); and describing variability between assessors (n = 2 papers reported data on interobserver variability for index test or reference standard).

Findings

The individual included studies have been described in detail in Characteristics of included studies and Table 1; we have also presented tabulated data for test accuracy by covariate (Summary of findings 2) and form of IQCODE threshold (Summary of findings 3). The total number of participants across the studies was 2644 (range: 37 to 684), of whom 379 (14%) had a clinical dementia diagnosis. The scope of the included studies was international; included data sets were from six countries (Australia, Canada, Japan, Spain, Sri Lanka and Thailand) (Appendix 11).

Certain papers contained more than one data set. For one paper the data sets were independent (one urban, one rural) and so we included these as separate entries (Morales 1997 (urban); Morales 1997 (rural)). One study had a single population assessed by two independent assessors (one with neurology training and one with psychiatry training). We used data from only one assessor for our analysis (Jorm 1996 (psychiatry), favouring the data closest to the expected population dementia prevalence.

Ten different versions of IQCODE were used in the included studies (Appendix 11) and eight different diagnostic thresholds (3.0, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 4.0) were used to define a positive IQCODE. We limited our analysis to the validated forms of IQCODE that are in common clinical use, that is the 26 and 16-item questionnaires.

Within the pre-specified thresholds chosen for analysis there was a spread of sensitivity and specificity (sensitivity range: 44% to 92%; specificity range: 55% to 96%).

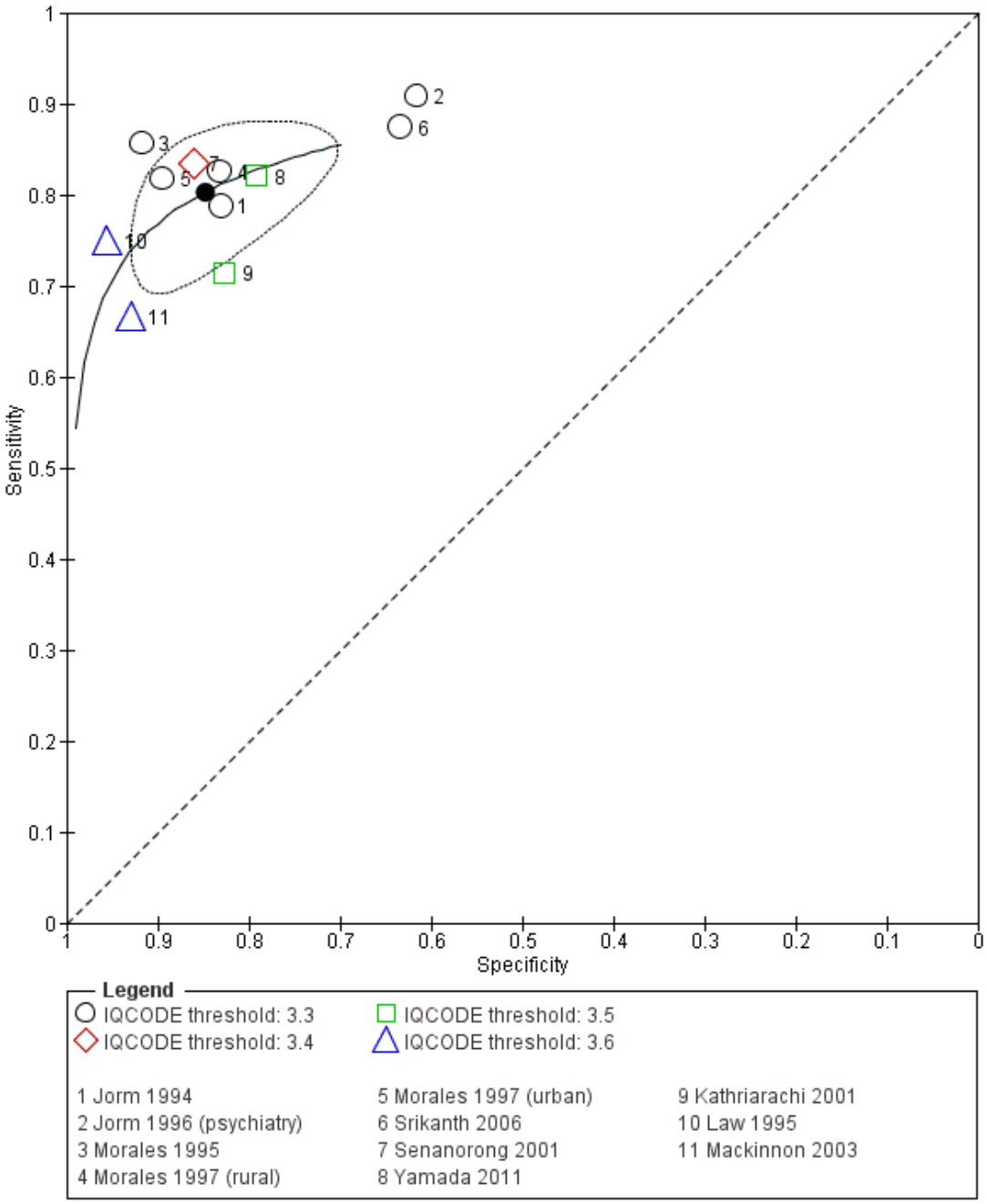
IQCODE (combined 16 and 26-item questionnaire)

Overview analysis - IQCODE using a 3.3 threshold or closest

Across 10 studies there were 11 data sets that contained relevant data (n = 2644). Sensitivity was 0.80 (95% CI 0.75 to 0.85); specificity 0.85 (95% CI 0.78 to 0.90). The overall positive likelihood ratio was 5.27 (95% CI 3.7 to 7.5) and the negative likelihood ratio was 0.23 (95% CI 0.19 to 0.29).

The summary ROC curve describing test accuracy across the included studies is presented in Figure 4.

Figure 4. Summary ROC Plot, IQCODE using a 3.3 threshold score or nearest. The dark point is a summary point, the broken line represents 95% CI



IQCODE 3.3 threshold or closest - comparing 26 and 16-item IQCODE

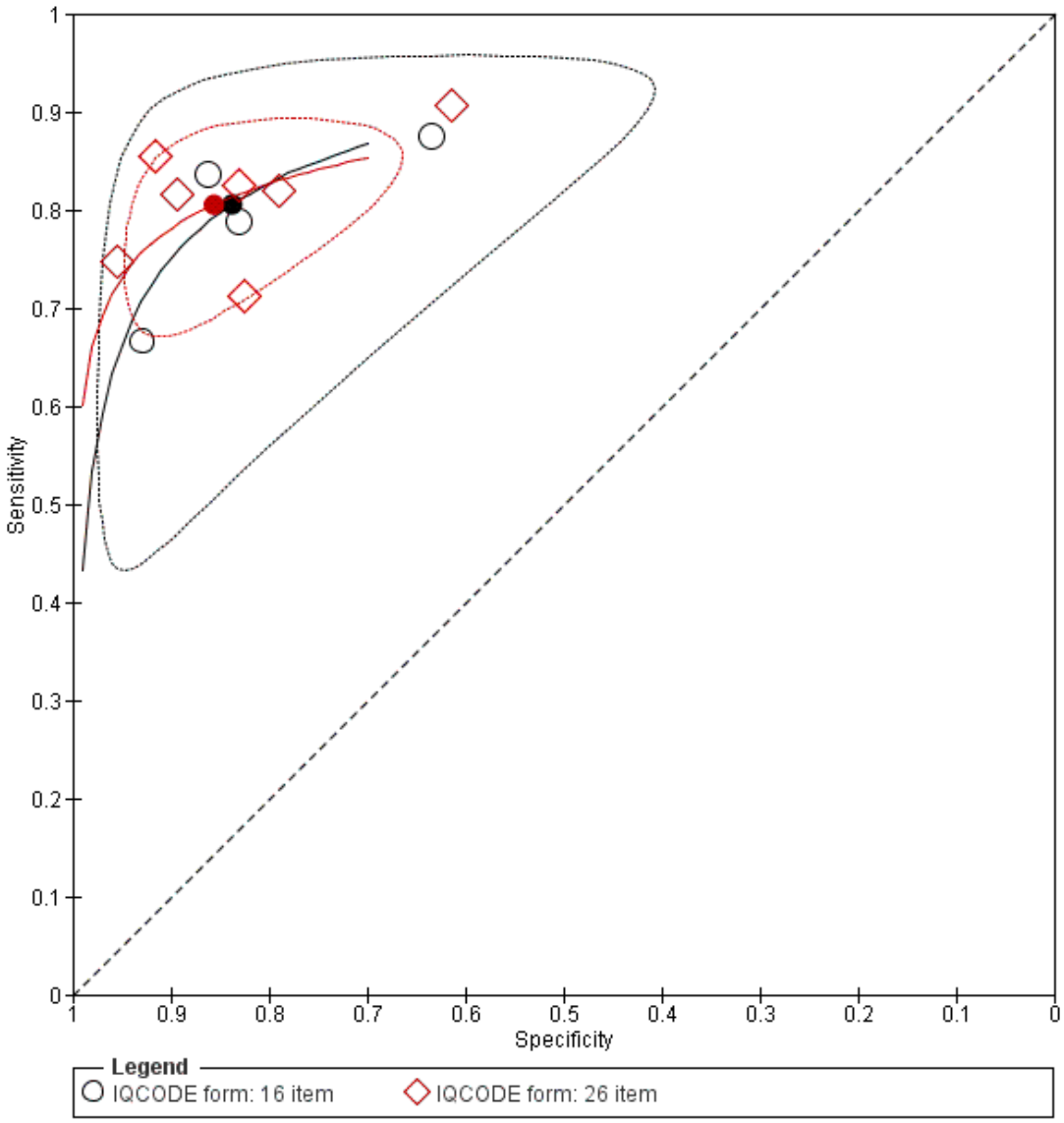
We used the overview data set to examine the effect of heterogeneity relating to IQCODE format (traditional 26-item or short form 16-item).

Analysis of the studies using the 26-item IQCODE (n = 7 data sets) gave sensitivity of 0.80 (95% CI 0.73 to 0.85); specificity 0.86 (95% CI 0.74 to 0.95). The overall positive likelihood ratio was 5.6 (95% CI 3.4 to 9.1) and the negative likelihood ratio was 0.24 (95% CI 0.18 to 0.31).

Analysis of studies using the 16-item IQCODE (n = 5 data sets) gave sensitivity of 0.80 (95% CI 0.74 to 0.85); specificity 0.82 (95% CI 0.70 to 0.89). The overall positive likelihood ratio was 4.2 (95% CI 2.6 to 6.8) and the negative likelihood ratio was 0.24 (95% CI 0.10 to 0.65).

Comparing the two there was no difference in accuracy, with a relative sensitivity of the 26-item versus 16-item IQCODE of 1.00 (95% CI 0.91 to 1.11) and relative specificity 0.94 (95% CI 0.82 to 1.09) (Figure 5).

Figure 5. Summary ROC plot of IQCODE 3.3 threshold or nearest, comparing short form (16 item) and traditional IQCODE.



As there was no difference we presented further data as the combined (26 and 16-item IQCODE together) test accuracy.

IQCODE 3.3 threshold or closest - comparing English and non-English language IQCODE

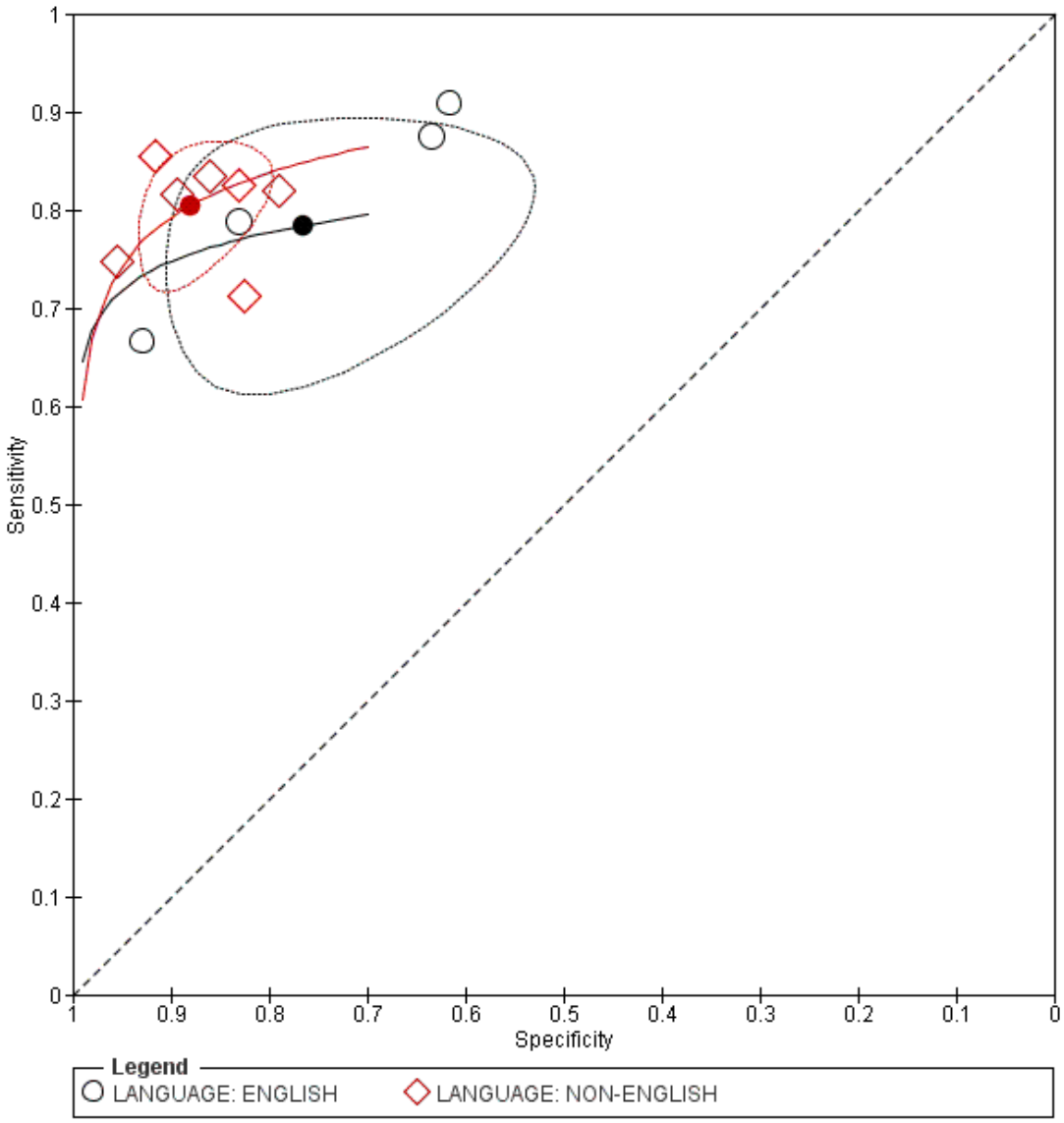
We coded the language of IQCODE administration as a covariate. Study numbers did not allow analysis by individual languages and so we compared the IQCODE in the original wording (English language) with all translated IQCODE forms (non-English language).

Analysis of studies using English language IQCODE (n = 5 data sets) gave sensitivity of 0.78 (95% CI 0.69 to 0.85); specificity 0.77 (95% CI 0.63 to 0.86). The overall positive likelihood ratio was 6.7 (95% CI 4.6 to 9.7) and the negative likelihood ratio was 0.28 (95% CI 0.20 to 0.41).

Analysis of studies using non-English language IQCODE (n = 7 data sets) gave sensitivity 0.80 (95% CI 0.74 to 0.85); specificity 0.88 (95% CI 0.82 to 0.92). The overall positive likelihood ratio was 6.7 (95% CI 4.6 to 9.7) and the negative likelihood ratio was 0.13 (95% CI 0.08 to 0.24).

Comparing the two there was no difference in accuracy, with a relative sensitivity of 1.03 (95% CI 0.91 to 1.16) and relative specificity of 1.15 (95% CI 0.98 to 1.34) (Figure 6).

Figure 6. Summary ROC Plot of pooled IQCODE data at a 3.3 threshold (or nearest value), with language as covariate. The dark point is a summary point, the broken line represents 95% CI



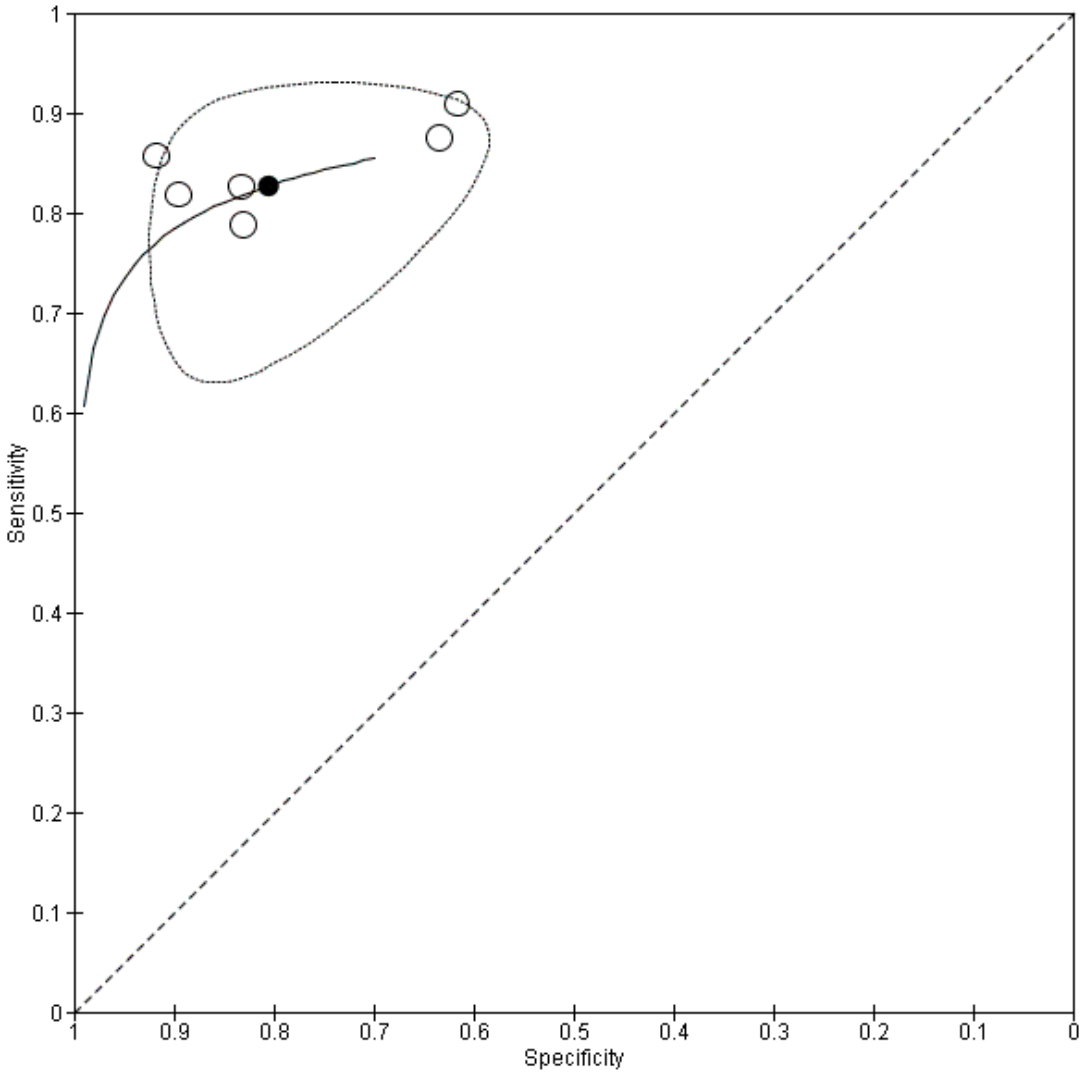
As there was no significant difference between groups we presented the data (all languages) for each of our pre-specified thresholds.

IQCODE test accuracy at differing diagnostic thresholds

We calculated test accuracy at our pre-specified IQCODE thresholds. We chose to present a summary ROC curve for those analyses with greater than three included studies.

IQCODE 3.3 threshold: there were six data sets* (n = 1232) that contained relevant data. The sensitivity was 0.83 (95% CI 0.74 to 0.90); specificity 0.80 (95% CI 0.70 to 0.88). The overall positive likelihood ratio was 4.25 (95% CI 2.75 to 6.56) and the negative likelihood ratio was 0.21 (95% CI 0.14 to 0.32) (Figure 7).

Figure 7. Summary ROC Plot of combined(16 and 26 item) IQCODE data using a 3.3 threshold score.The dark point is a summary point, the broken line represents 95% CI



IQCODE 3.4 threshold: there were three data sets* (n = 988) that contained relevant data. The sensitivity was 0.84 (95% CI 0.70 to 0.93); specificity 0.80 (95% CI 0.65 to 0.90). The overall positive likelihood ratio was 4.42 (95% CI 2.47 to 7.90); the negative likelihood ratio was 0.19 (95% CI 0.10 to 0.35).

IQCODE 3.5 threshold: there were three data sets* (n = 1144) that contained relevant data. Sensitivity was 0.82 (95% CI 0.75 to 0.87); specificity 0.84 (95% CI 0.80 to 0.88). The overall positive likelihood ratio was 5.09 (95% CI 4.08 to 6.33); the negative likelihood ratio 0.22 (95% CI 0.16 to 0.29).

IQCODE 3.6 threshold: there were three studies (n = 1215) that contained relevant data. Sensitivity was 0.78 (95% CI 0.68 to 0.86); specificity was 0.87 (95% CI 0.71 to 0.95). The overall positive likelihood ratio was 6.00 (95% CI 2.72 to 13.26); the negative likelihood ratio was 0.25 (95% CI 0.18 to 0.34).

** Certain papers included more than one data set*

Heterogeneity relating to dementia diagnosis

A quantitative analysis of the effect of dementia diagnosis criteria (reference standard) was not possible as all but one (Jorm 1996

(psychiatry) of the studies that specified the approach to dementia diagnosis used the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (DSM) to define the dementia state. The remaining study used the World Health Organization (WHO) International Statistical Classification of Diseases and Related Health Problems (ICD) for diagnosis. This study also compared neurologists' and psychiatrists' diagnoses against the IQCODE and found that IQCODE was more sensitive compared to a psychiatrist's diagnosis of dementia (Jorm 1996 (psychiatry). A further original aim was to describe the accuracy of the IQCODE for diagnosis of Alzheimer's disease dementia. We were unable to assess this based on the available data as only one study defined specific dementia diagnoses (Law 1995); all other studies described the accuracy of IQCODE for diagnosis of all cause dementia only.

Other sources of heterogeneity and sensitivity analyses

One study was of community-dwelling stroke-survivors (Srikanth 2006); we performed a sensitivity analysis by removing these data from our pooled estimates. We found little difference in test accuracy when this study was removed (at a 3.3 threshold, sensitivity 0.81, 95% CI 0.73 to 0.87; specificity 0.81, 95% CI 0.70 to 0.85). We performed a sensitivity analysis removing studies with a 'young' population. Not all studies provided data on the age of participants and descriptive metrics differed across the papers (Table 2). Two authors (TJQ, PF) reviewed the ages of the included populations and concluded that one study contained a 'younger' cohort likely to meet our pre-specified arbitrary cut-off of more than 20% aged less than 65 years (Senanorong 2001; see Table 2). This study also had an unusually high prevalence of dementia suggesting that a case-control methodology was employed, although this was not explicitly stated in the paper. We performed a sensitivity analysis excluding this study. Test accuracies at thresholds of 3.4 and 3.5 were similar after exclusion of this study (sensitivity 0.82, 95% CI 0.72 to 0.89; specificity 0.79, 95% CI 0.66 to 0.89 at a 3.4 cut-point; sensitivity 0.82, 95% CI 0.74 to 0.88; specificity 0.83, 95% CI 0.78 to 0.87 at a 3.5 cut-point).

Given the modest numbers of papers and the clinical heterogeneity we did not perform any further sensitivity analysis by QUADAS-2 metrics or other factors.

Summary of findings

Study ID	Country	Subjects (n)	IQCODE version	Language	Dementia diagnosis	Dementia prevalence N (%)		Other assessments
Jorm 1994	Australia	684	26 & 16 item	English	DSM IIIr	n=52	8%	-
Jorm 1996 (psychiatry)	Australia	144	26 & 16 item	English	ICD9	n=11	8%	MMSE
Kathriarachi 2001	Sri Lanka	37	26 item	Sinhalese	‘ ‘ clinical assessment”	n=14	38%	MMSE, CDR
Law 1995	Canada	237	26 item	French	DSM IIIr	n=32	14%	MMSE
Mackinnon 2003	Australia	646	16 item	English	DSM IIIr	n=36	6%	MMSE
Morales 1995	Spain	68	26 & 17 item	Spanish	DSM IIIr	n=7	10%	MMSE
Morales 1997 (rural)	Spain	160	26 item	Spanish	DSM IIIr	n=23	14%	MMSE
Morales 1997 (urban)	Spain	97	26 item	Spanish	DSM IIIr	n=11	11%	MMSE
Senanorong 2001	Thailand	160	16 & 3 item	Thai	DSM IV	n=73	46%	TMSE
Srikanth 2006	Australia	79	16 item	English	DSM IV	n=8	10%	S-MMSE
Yamada 2011	Japan	423	26 item	Japanese	DSM IV	n=112	26%	

See Characteristics of included studies for more detailed study descriptors.
Abbreviations: DSM - American Psychiatric Associationm Diagnostic and Statistical Manual of Mental Disorders; MMSE - Mini Mental State Examination; AMT - Abbreviated Mental Test; CDR - Clinical Dementia Rating Scale; TMSE - Thai Mental State Exam; S-MMSE - standardised Mini Mental State Examination.

4 What is the accuracy of the Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE) test for detection of dementia using different versions of IQCODE and using different languages of administration				
Population	Community-dwelling older adults, with no restrictions placed on case-mix of included cohort			
Setting	'Community' setting; this setting was intended to represent a population screening context. Many of the included studies, although fulfilling our pre-specified inclusion criteria, were of selected population groups (for example stroke-survivors; ex-prisoners of war) the effect of these studies is described in the 'heterogeneity' section of results			
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 a commonly used short form IQCODE with 16 items			
Reference Standard	Clinical diagnosis of dementia made using any recognised classification system			
Studies	Cross-sectional studies were included, we did not include case-control studies			
Comparative analyses				
Test	No. of participants (studies)	Dementia prevalence total across studies	Findings	Implications
26 item versus 16 item IQCODE	Total: n=2644 (10 studies, 11 data sets) 26 item n=1075 (7 studies, 8 datasets)	Total n=379 (14%) 26 item n=210 (20%) 16 item n=169 (11%)	No difference in accuracy. Relative sensitivity of 26-item versus 16-item IQCODE: 1.00 (95% CI 0.91 to 1.11) Relative specificity of 26 item versus 16-item IQCODE: 0.94 (95% CI 0.82 to 1.09)	Short form IQCODE may be preferred as lesser test burden with similar accuracy
English language versus non-English	Total: n=2644 (10 studies, 11 data sets) English: n=1553 (4 studies)	Total: n=379 (14%) English: n=107 (6%) Non-English: n=272 (25%)	No significant difference in accuracy. Relative sensitivity of English language versus non-English language: 1.03 (95% CI 0.91 to 1.16) Relative specificity of English language versus non-English language: 1.15 (95% CI 0.98 to 1.34)	IQCODE accuracy is not substantially influenced by language of administration
CAUTION: The results on this table should not be interpreted in isolation from the results of the individual included studies contributing to each summary test accuracy measure. These are reported in the main body of the text of the review				

What is the accuracy of the Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE) test for detection of dementia when differing thresholds are used to define IQCODE positive cases	
Population	Community-dwelling older adults, with no restrictions placed on case-mix of included cohort
Setting	'Community' setting; this setting was intended to represent a population screening context. Many of the included studies, although fulfilling our pre-specified inclusion criteria, were of selected population groups (for example stroke-survivors; ex-prisoners of war) the effect of these studies is described in the 'heterogeneity' section of results
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 a commonly used short form IQCODE with 16 items
Reference Standard	Clinical diagnosis of dementia made using any recognised classification system
Studies	Cross-sectional studies were included, we did not include case-control studies

Test	Summary accuracy (95%CI)	No. of participants (studies)	Dementia prevalence	Implications Quality and comments
IQCODE cut-off 3.3 or nearest	sensitivity: 0.80 (95% CI 0.75 to 0.85) specificity: 0.85 (95% CI 0.78 to 0.90) positive LR: 5.27 (95% CI 3.70 to 7.50) negative LR: 0.23 (95% CI 0.19 to 0.29)	n=2644 (10 studies, 11 datasets)	n=379 (14%)	There is no obvious preferred cut-off for IQCODE accuracy, within the threshold values commonly used in clinical practice and research So we focus on the summary data across all cut-points. The dementia prevalence across studies is higher than would be expected for this population. Using the accuracy figures re-calculating for a typical population; In the UK 9.9 million people are aged over 65, current estimates are of around 6.6% dementia prevalence. At the IQCODE accuracy calculated, using IQCODE alone to ' ' screen" for dementia would result in:

				87,120 people with dementia not being picked up and 1,314,660 dementia free people being given a possible diagnosis of dementia
IQCODE cut-off 3.3	sensitivity: 0.83 (95% CI 0.74 to 0.90) specificity: 0.80 (95% CI 0.70 to 0.88) positive LR: 4.25 (95% CI 2.75 to 6.56) negative LR: 0.21 (95% CI 0.14 to 0.32)	n=1232 (5 studies, 6 datasets)	n=112 (9%)	
IQCODE cut-off 3.4	sensitivity: 0.84 (95% CI 0.70 to 0.93) specificity: 0.80 (95% CI 0.65 to 0.90) positive LR: 4.25 (95% CI 2.47 to 7.90) negative LR: 0.19 (95% CI 0.10 to 0.35)	n=988 (3 studies)	n=136 (14%)	
IQCODE cut-off 3.5	sensitivity: 0.82 (95% CI 0.75 to 0.87) specificity: 0.84 (95% CI 0.80 to 0.88) positive LR: 5.09 (95% CI 4.08 to 6.33) negative LR: 0.22 (95% CI 0.16 to 0.29)	n=1144 (3 studies)	n=178 (16%)	
IQCODE cut-off 3.6	sensitivity: 0.78 (95% CI 0.68 to 0.86) specificity: 0.87 (95% CI 0.71 to 0.95) positive LR: 6.00 (95% CI 2.72 to 13.26) negative LR: 0.25 (95% CI 0.18 to 0.34)	n=1215 (3 studies)	n=180 (15%)	

CAUTION: The results on this table should not be interpreted in isolation from the results of the individual included studies contributing to each summary test accuracy measure. These are reported in the main body of the text of the review

DISCUSSION

Summary of main results

We offer a synthesis of the published data describing the accuracy of the IQCODE questionnaire tool for detection of dementia within community-dwelling populations.

Our results suggests that although the IQCODE has reasonable test properties, for example the positive likelihood ratio of around 5 and negative likelihood ratio of around 0.2 are classically interpreted as indicative of a 'moderately good test', the test alone may not be suited for dementia screening within community dwelling older adults.

For a clinical assessment the preferred pattern of diagnostic test accuracy (DTA), optimising sensitivity versus optimising specificity, will vary with the purpose of the test. The utility and limitations of screening all community-dwelling older adults for cognitive problems is a topic that is attracting considerable international debate. Our data show that even for a 'good' initial assessment like IQCODE, at a population level the number of false positives and false negatives is still considerable. Applying our summary data to a population such as the UK, where 9.2 million adults are aged over 65 and 6.6% (435,600) of this group may have dementia, we see that even modest problems in test accuracy can be associated with considerable numbers of false diagnoses or false reassurance at population level (using these population numbers and at sensitivity of 0.80 / specificity of 0.84 we find false positive numbers of around 1314,660; false negative numbers of around 87,120). We appreciate that in practice the use of such tests is more pragmatic, but we give this example to illustrate the potential effects of IQCODE screening at a population level.

Accepting that IQCODE is a reasonable initial test, albeit is perhaps not sufficient as a single screening test, our pre-specified analyses around heterogeneity were designed to provide guidance on optimal IQCODE administration with specific reference to form of IQCODE; language of IQCODE and preferred test positive cut-point.

There was little difference in sensitivity across the predefined diagnostic cut-points. We had expected a more pronounced 'trade-off' between sensitivity and specificity at differing thresholds. Possible explanations are that the thresholds are too close together to see differences in accuracy between neighbouring cutoffs or that any differences are lost in between study heterogeneity. We can conclude that at the IQCODE values commonly described in research, there is little to choose between the thresholds. There was a suggestion that sensitivity began to fall at cut-points above 3.5 and a trend towards improved specificity with increasing cut-point from 3.3 to 3.6. It would seem intuitive that scores above and below these values would have a more marked difference in sensitivity to specificity ratio. In certain situations, for example in dementia screening where specificity may be preferred to avoid false positive diagnosis, a cutoffs below 3.3 may be preferred. However, we found few published studies describing thresholds less than 3.3 or greater than 3.6 and so at present this hypothesis is speculative.

There were many differing forms of IQCODE application described across the included papers. We pre-specified a comparative analysis of IQCODE when used with the traditional 26 questions and a short form with 16 questions. As the tools had similar accuracy we believe pooling data across these two IQCODE formats was valid. There were insufficient data to describe accuracy of IQCODE assessments that did not use the standard 26 or 16 item questionnaires and we were wary of describing test accuracy of unvalidated IQCODE based assessments.

The other area of heterogeneity in IQCODE application was for language of administration. There were insufficient numbers of papers to allow a valid analysis of the effect of individual languages of IQCODE, however summary analysis using dichotomised language ('English' or 'non-English') as a covariate suggested no significant difference. Although not reaching significance, there was a trend towards differing accuracy. The effect was not as expected with the non-English language IQCODE seemingly having improved accuracy. However differences were modest and it seems likely that some of these difference will relate to differing study methodologies and populations rather than the scale itself. Nonetheless we should be mindful of potential language effects in interpreting the pooled analysis and future studies should detail the language(s) of administration of tests employed.

We restricted our analysis to the healthcare setting of "community based studies". This setting and terminology was chosen prior to searching and review of the literature. Our intention was to assess those studies where participants had not been included on the basis of cognitive testing or symptoms and we suspected that included studies would have a population level assessment methodology. Across the literature describing IQCODE, the "community" setting proved difficult to operationalize and included a number of differing population sampling methods and study types. Certain included studies could be criticised for not conforming to usual definitions of unselected, community dwelling older adults (for example one study was of ex-servicemen only). We included all community based studies if participants were not selected on the basis of a factor that may relate to dementia or cognitive functioning. One study, although community based, included stroke survivors only. Clearly this group may differ from a non-stroke population and we explored this using sensitivity analyses. In fact, the test accuracy of IQCODE was similar comparing stroke and non-stroke, albeit confidence intervals were necessarily larger.

Even where papers seemed to have a population based sampling frame, the prevalence of dementia was unexpectedly high and we must be cautious in our interpretation of these data. For unselected community assessment we would expect a prevalence of dementia in keeping with previous population estimates (5% of adults age over 60 years; 6% to 7% of adults aged over 65 years). Only one of our included papers (Mackinnon 2003) had proportions

with dementia in this range. One study had a younger population and a high prevalence of dementia suggesting a case-control methodology,albeit this was not explicit in the manuscript, again we explored the effect of removing this potentially unrepresentative study with sensitivity analyses and found little difference to pooled estimates with exclusion of the paper.

In many of the included studies there was substantial potential for bias and reporting quality was suboptimal. In general, authors gave sufficient detail and were robust in their clinical dementia assessment (reference standard); however methodology and reporting of patient sampling and use of IQCODE could be improved. Assessment of quality is dependent on adequate reporting and there were many examples where QUADAS scoring was complicated by insufficient detail. One example is the blinding of dementia assessors to IQCODE data, particularly as dementia diagnosis is often partly predicated on information from informants. We hope that the proposed dementia specific reporting guidance of STARDdem may improve quality in future studies that use dementia as reference standard.

Strengths and weaknesses of the review

The strength of this review was its focused study question and setting. This review was limited to studies of community dwelling adults. While this approach avoids heterogeneity that may be introduced by test 'setting' it does limit applicability to other settings and we should not extrapolate the data presented in this review to hospital or primary care populations. Reviews of the test accuracy of IQCODE in other settings and of the test accuracy of other direct and informant tests are planned as separate Cochrane reviews. We performed a comprehensive and sensitive literature search, encompassing cross-disciplinary electronic databases and test accuracy specific resources. Our primary search was complemented by contact with other authors working in the field and we are grateful for all the helpful responses we received. We did not limit by language of paper and this proved to be important as studies of IQCODE were international and several papers required translation. An unexpected finding was the modest numbers of studies describing IQCODE accuracy in community settings from United Kingdom and North America.

Due to the modest numbers of papers, we pooled data for our summary analysis across various forms of IQCODE. Our comparative analysis would suggest that 16 and 26 item IQCODE have similar test accuracy, however language of administration may influence properties and as a result our summary data must be interpreted with some caution.

We endeavoured to be as robust as possible in our assessment of included studies. Our approach to risk of bias assessment was informed by a short life working group that met to define relevant and workable anchoring statements and definitions of criteria (Appendix 9). As we felt that assessment of quality should include a measure of quality of reporting we also assessed the included papers using the STARD approach (Appendix 7).

Important clinical and demographic details that could impact on the interpretation of our IQCODE data were not consistently reported and we could not describe the effect of factors such as nature of the informant and severity of dementia. For translating test accuracy studies to clinical practice, an approach that describes numbers who could not be tested with index and reference standard is often useful (i.e. an intention to diagnose approach with a two by three or three by three table; rather than the standard two by two table) (Schuetz 2012). We did not collect data in this format and at present we do not have techniques to allow pooling of such data.

We await the results of ongoing systematic reviews and meta-analyses of other dementia assessment strategies (many of which are being completed by the Cochrane Dementia and Cognitive Improvememt Group) before we can begin to compare assessments and suggest the optimal tests for a particular patient group or clinical indication.

Applicability of findings to the review question

We found studies relevant to our focused study question and were able to give summary estimates for certain of our pre-specified co-variates of interest.. Our primary objective was to determine the diagnostic accuracy of the informant based questionnaire IQCODE, for detection of all cause (undifferentiated) dementia in community-dwelling adults and we provide summary data that hopefully will help clinicians and policy makers understand the properties of IQCODE as an initial assessment in this setting.

A priori we had defined a number of subgroup and sensitivity analyses, the limited number of included papers precluded many of these analyses and we have not definitively answered our secondary questions of describing the effect of age or dementia diagnosis on test accuracy metrics. Further potential heterogeneity will be introduced by the "stage"/severity of dementia at time of diagnosis, as diagnosis will be easier in advanced disease than in early disease. No included studies gave data on severity of diagnosis that would allow us to describe this effect and so based on available data we can make no specific comment on use of IQCODE in, for example, early stage dementia.

When planning the analysis we had conceptualised the "community" setting as being closest to unselected, population screening. Most of the included papers, while fulfilling our criteria for "community", still described selected populations. It may be that this is of only minor consequence, as test accuracy of IQCODE was similar even when comparing highly selected groups (for example stroke survivors) to the pooled result.

AUTHORS' CONCLUSIONS

Implications for practice

Accepting the limitations of included studies and the potential for baises, results were fairly consistent across the studies and allow us to give some guidance on the use of IQCODE. Published data suggest that for initial assessment of dementia in older adults the IQCODE with cut-points of 3.3 to 3.6 could be used, however as a single assessment tool IQCODE properties may not be suited to population level screening. We extrapolated the IQCODE summary accuracy data to a United Kingdom context as exemplar and can see that using IQCODE exclusively will lead to substantial false positive diagnosis. Given the public perception of dementia, it is arguable that the distress caused by assigning a dementia label to a person without the disease is greater than the potential harm of initially missing dementia on screening. These are important concepts that need to be considered if large scale cognitive screening is to be introduced.

The choice of a screening tool or triage tool for cognitive assessment will not only be driven by test accuracy. Strengths of the IQCODE from a clinician or healthcare perspective are that it is copyright free; available in many languages; and relatively quick and easy to complete. In general, informant-based assessments that do not rely on direct patient testing can capture change over time and are less prone to social-cultural biases (Jorm 2000A; Larner 2013). These are all factors that make IQCODE attractive as an initial assessment tool and explain why it is popular in its clinical and research use.

Our analyses of heterogeneity suggest that 26 and 16-item IQCODE have similar test accuracy. It would seem sensible to recommend the short version of IQCODE as administration time or burden is less with comparable accuracy. Other short form versions of IQCODE have been described, at present there are insufficient data available to recommend use of these other modified IQCODE formats.

There was a trend, albeit not reaching significance, to suggest that the language of administration may impact on IQCODE accuracy. Our findings do not imply that certain languages of assessment are more or less accurate. The safest interpretation of these data is as a reminder that translating IQCODE items to other languages needs to be sensitive to idioms and cultural nuances. We would encourage assessors using a non-English language IQCODE to ensure that any translation and validation process has been suitably robust.

As a single test review, our data do not allow us to comment on how the IQCODE performs in relation to other tests. Given the large number of assessment tools potentially available, this is the question that may be of most interest to clinicians. In many papers and in clinical practice, the information from IQCODE is often used in tandem with a direct patient cognitive assessment tool such as MMSE. While combining instruments is intuitively attractive, at present we do not have a systematic review of the properties of this approach.

Implications for research

Our pooled analysis gives a large test population and the associated estimates of diagnostic accuracy are reasonably robust. However, we still encourage further study of the properties of IQCODE. As an example, despite our focused study question around community setting, the included studies in our review were largely not typical of an unselected, older adult population. The ideal study would involve stratified sampling and testing based, for example, on census data, such approaches have been used in seminal work describing the epidemiology of dementia (Ferri 2005).

IQCODE test accuracy was maintained comparing 26 and 16-item formats. IQCODE versions with even fewer than 16 items have been described, although numbers were too small for pooled analysis in this review. In practice a brief assessment tool is an attractive option if diagnostic accuracy can be maintained. We would recommend further study of shortened (less than 16-item) IQCODE properties.

Our review had a deliberately focused agenda and our data do not allow us to extrapolate the diagnostic properties of IQCODE to other healthcare settings. We recognise that dementia assessment with additional informant interview is common in primary care and hospital settings and reviews of the IQCODE when used in these settings are now required. We have alluded to the need for comparative studies of various tools used alone or in combination. The ongoing body of work by the Cochrane group describing the test accuracy of commonly used direct and indirect tests will offer a substrate for future indirect comparative meta-analysis.

Our assessments of reporting quality and risk of bias are concerning but in keeping with results from other areas of dementia research. We urge dementia researchers to work towards improved consistency in both methodology and reporting to assist future reviews of the diagnostic accuracy of tests. The use of dementia-specific guidance such as the proposed STARDdem initiative may assist future trialists.

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 D Salmon, Dr S Sikkes, Dr G Potter, Dr M Razavi, Prof H Henon, Dr JFM de Jonghe, Dr V Isella, Dr AJ Larner, Dr B Rovner, Dr M Krogseth.

We thank Dr Y Takwoingi for assistance with the statistical analysis.

REFERENCES

References to studies included in this review

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 Medicine* 1994;**24**: 145–53.

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

Kathriarachi 2001 *{published and unpublished data}*
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.

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

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.

Morales 1995 *{published and unpublished data}*
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's Disease and Associated Disorders* 1995;**9**:105–11.

Morales 1997 (rural) *{published and unpublished data}*
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 1997 (urban) *{published and unpublished data}*
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.

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

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**:295–305.

Yamada 2011 *{published and unpublished data}*
Mimori Y, Miyachi T, Ohshita T, Nakamura S, Yamada M, Sasaki H, Suzuki G. Cognitive decline and detection of dementia among the Japanese population. Anlaysis with the CASI and IQCODE. conference proceedings (poster). 2011.

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 Pisquiatria* 2008;**30**:346–9.

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.

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.

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. *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.

Diesfeldt 2007 *{published data only}*
Diesfeldt HFA. Discrepancies between IQCODE and cognitive test performance. *Tijdschrift voor Gerontologie en Geriatrie* 2007;**38**:199–209.

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.

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 (abstract) *{published data only}*
Finelli L, Kunze U, Gautier A, Gomez-Mancilla B, Monsch AU. Algorithms to retrospectively diagnose mild cognitive impairment and dementia in a longitudinal study of ageing and dementia. abstract ICAD. 2009.

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.

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.

Henon 2001 *{published data only}*
Henon H, Durieu I, Guerouaou D, Lebert F, Pasquier F, Leys D. Poststroke dementia incidence and relationship to pre-stroke cognitive decline. *Neurology* 2001;**57**:1216–22.

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–s80.

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

Jorm 1989 *{published data only}*
Jjorm AF, Scott R, Jacomb PA. Assessment of cognitive decline in dementia by informant questionnaire. *International Journal of Geriatric Psychiatry* 1989;**4**:35–9.

Jorm 1989b *{published data only}*
Jorm AF, Jacomb PA. The 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 1996 (Age and Ageing *{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. *Alzheimers 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 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.

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, Ames D. The combination of cognitive testing and an informant questionnaire in screening for dementia. *Age and Ageing* 2003;**32**:541–547.

Krogseth 2011 *{published data only}*
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.

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.

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 Geriatric Psychiatry and Neurology* 2012;**25**:227–32.

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.

Mimori (abstract) ***{published data only}***
Mimori Y. Cognitive decline and detection of dementia among the Japanese population analysis with CASI and IQCODE. abstract. 2010:s451.

Morales-Gonzalez 1992 ***{published data only}***
Morales-Gonzalez JM, Gonzalez-Montalvo JL, De Ser Quijano T, Bermejo Pareja F. Validation of the S-IQCODE [Original paper in Spanish]. *Archivos de Neurobiologiá (Madrid)* 1992;**55**: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. Improving 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, 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 & Dementia* 2009;**5**:445–53.

Razavi 2011 ***{published and unpublished data}***
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's Disease and Associated Disorders* 2012;**26**:44–9.

Sanchez 2009 ***{published data only}***
dos Santos Sanchez MA, Lourenco RA. IQCODE cross cultural adaptation for use in Brazil. *Cadernos de Saude Publica* 2009;**25**:1455–65.

Schofield 2006 ***{published data only}***
Schoffield 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.

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 Tropical Medicine and Public Health* 2006;**37**:587–94.

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**:214–20.

Tang 2003 ***{published data only}***
Tang WK, Sandra SMC, Chiu HFK, Wong KS, Kwok TCY, Mok V, Ungvari GS. 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 PL. Utility of the IQCODE for dementia in a Japanese American population. *Hawai'i Medical Journal* 2006;**65**: 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. abstract (World Congress of Neurology). 1997.

Wolfe 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;**10**:1178–80.

Zevallos-Bustamente 2003 *{published data only}*
Zevallos Bustamante SE, Bottino CMC, Lopes MA, Dionísio Azevedo D, Hototatian SR, Litvoc J, Filho JW. 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, Wang L. Telephone IQCODE for dementia. abstract. 2004.

Additional references

Birks 2006
Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database of Systematic Reviews* 2006;**1**:CD005593.

Bossuyt 2003
Bossuyt PM, Reitsma JB, Bruns DE Gatsonis CA, Glasziou PP, Irwig LM, et al.Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;**326**:41–4.

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**:927–37.

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

Brunet 2012
Brunet MD, McCartney M, Heath I, Tomlinson J, Cosgrove J, Deveson P, et al.Open letter to the Prime Minister and Chief Medical Officer for England. There is no evidence base for proposed dementia screening. *BMJ* 2012;**345**:e8588.

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**:1051–9.

Clare 2003
Clare L, Woods B. Cognitive rehabilitation and cognitive training for early-stage Alzheimer's disease and vascular dementia. *Cochrane Database of Systematic Reviews* 2003;**4**: CD003260.

Cordell 2013
Cordell CB, Borson B, Boustani M, Chodosh J, Reuben D, Verghese J, et al.Medicare Detection of Cognitive Impairment Group. Alzheimer's Associaton recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimer's & Dementia* 2013;**9**:141–50.

Cordoliani-Mackowiak 2003
Cordoliani-Mackowiak MA, Henon H, Pruvo JP, Pasquier F, Leys D. Poststroke dementia influence of hippocampal atrophy. *Archives of Neurology* 2003;**60**:585–90.

Cullen 2007
Cullen B, O'Neill B, Evans JJ, Coen RF, Lawlor BA. A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery, and Psychiatry* 2007;**78**:790-9.

Erkinjuntti 2000
Erkinjuntti T, Inzitari D, Pantoni L. Research criteria for subcortical vascular dementia in clinical trials. *Journal of Neural Transmission. Supplementa* 2000;**59**:23-30.

Ferri 2005
Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, et al.Alzheimer's Disease International. Global prevalence of dementia: a Delphi consensus study. *Lancet* 2005;**366**: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**: 189-98.

Galvin 2005
Galvin JE. A brief informant interview to detect dementia. *Neurology* 2005;**65**:559–64.

Greenhalgh 2005
Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005;**331**:1064–5.

Hebert 2003
Hebert LE, Scherr PA, Bienas JL, Bennett DA, Evans DA. Alzheimers disease in the US population: prevalence

estimates using the 2000 census. *Archives of Neurology* 2003; **60**:1119–22.

Holsinger 2007
Holsinger T, Deveau J, Boustani M, Willimas JW. Does this patient have dementia. *JAMA* 2007;**21**:2391–404.

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

Jorm 1989A
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 2000A
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.

Matthews 2013
Matthews FE, Arthur A, Barnes LE, Bond J, Jagger C, Robinson L, Brayne C, Medical Research Council Cognitive Function and Ageing Collaboration. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *Lancet* 2013;**382**(9902):1405–12.

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**: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**: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.

McShane 2006
McShane R, Areosa Sastre A, Minakaran N. Memantine for dementia. *Cochrane Database of Systematic Reviews* 2006;**2**: CD003154.

Noel-Storr 2012
Noel-Storr AH, Flicker L, Ritchie CW, Nquyen GH, Gupta T, Wood P, et al.Systematic review of the body of evidence for the use of biomarkers in diagnosis of dementia. *Alzheimer's & Dementia* 2013;**9**(3):e96-e105. doi:10.1016/j.jalz.2012.01.014.

Reitsma 2005
Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**:982–90.

RevMan 2011
The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan) 5.1. Copenhagan: The Nordic Cochrane Centre, 2011.

Rockwood 1998
Rockwood K. Retrospective diagnosis of dementia using an informant interview based on the Brief Cgnitive Rating Scale. *International Psychogeriatrics* 1998;**10**:53–60.

Roman 1993
Roman GC, Tatemichi TK, Erkinjutti T, Cummings JL, Masdeu JC, Garcia JH, et al.Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology* 1993;**43**:250–60.

Savva 2009
Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C, et al.Age, neuropathology and dementia. *New England Journal of Medicine* 2009;**360**:230–9.

Schuetz 2012
Schuetz GM, Schlattmann P, Dewey M. USeo f 3x2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT-angiography studies. *BMJ* 2013;**345**: e6717.

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**:2964–8.

Yamada 2008
Yamada M, Mimori Y, Kasagi F, Miyachi T, Ohshita T, Sudoh S, et al.Incidence of dementia, Alzheimers disease and vascular dementia in a Japanese population radiation effects Research Foundation Adult Health Study. *Neuroepidemiology* 2008;**30**:152–60.

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

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

Jorm 1994

Study characteristics			
Patient sampling	Community sampling (unspecified), enriched with care-home residents		
Patient characteristics and setting	Community (n=945) and care-home dwelling older adults (n=100) approached; n=684 included Community setting		
Index tests	IQCODE 16 and 26 item, English language		
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IIIr, informed by the Canberra Interview for the Elderly		
Flow and timing	Of 1045 potential subjects, 769 had an informant; of this group a clinical diagnosis was possible in 684. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
			High
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

If a threshold was used, was it pre-specified?	No		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	Unclear		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	Yes		
			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Were missing IQCODE results or un-interpretable IQCODE results reported	Yes		

Jorm 1996 (psychiatry)

Study characteristics			
Patient sampling	Subjects were ex-servicemen enrolled in a separate prospective study		
Patient characteristics and setting	Community-dwelling ex-servicemen (n=144) Community setting		
Index tests	IQCODE 16 and 26 items, English language		
Target condition and reference standard(s)	Clinical dementia diagnosis using ICD9		
Flow and timing	Of 209 potential subjects,144 had an informant and were included. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment		
Comparative			
Notes	These subjects were assessed by a psychiatrist		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
			High
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Were sufficient details given on IQCODE application for the test to be repeated in an inde-	Unclear		

Jorm 1996 (psychiatry) (*Continued*)

pendent study			
			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	No		
			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Were missing IQCODE results or un-interpretable IQCODE results reported	Unclear		

Kathriarachi 2001

Study characteristics	
Patient sampling	Stratified community sampling, using census data and door to door assessment in a semi-urban setting
Patient characteristics and setting	Community-dwelling older adults (n=37) Community setting

Index tests	IQCODE 26 item, Sinhalese language		
Target condition and reference standard(s)	Clinical diagnosis of dementia following psychiatrist's review		
Flow and timing	Of 1400 potential subjects,40 were "randomly" selected for assessment and 37 assessed. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment		
Comparative			
Notes	Low numbers included and high prevalence of dementia		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			High
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	No		
			Unclear
DOMAIN 3: Reference Standard			

Kathriarachi 2001 (*Continued*)

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	No		
			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	No		
Were missing IQCODE results or un-interpretable IQCODE results reported	No		

Law 1995

Study characteristics	
Patient sampling	Age stratified sample of all community residents
Patient characteristics and setting	Randomly selected community-dwelling adults (n=237) Community setting
Index tests	IQCODE 26 item, French language
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IIIr

Flow and timing	Of 1800 potential subjects, 454 had psychiatric assessment of this group 364 had suitable informants and 237 were included. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment		
Comparative			
Notes			

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			Unclear
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	Yes		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Law 1995 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	Yes		
			Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Were missing IQCODE results or un-interpretable IQCODE results reported	Yes		

Mackinnon 2003

Study characteristics	
Patient sampling	Probability sample of older adults (age > 70 years) drawn from electoral data
Patient characteristics and setting	Probability sampling of community cohort (n=646) Community setting
Index tests	IQCODE 26 and 16 item, English language
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IIIr
Flow and timing	Of 945 potential subjects,694 had an informant and 646 were included. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment
Comparative	

Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	Yes		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Were sufficient details of dementia diagnostics given for the	Yes		

Mackinnon 2003 (*Continued*)

assessment to be repeated in an independent sample			
			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Were missing IQCODE results or un-interpretable IQCODE results reported	Unclear		

Morales 1995

Study characteristics	
Patient sampling	Random selection of community-dwelling older adults (age > 65) from census data with initial door to door assessment
Patient characteristics and setting	Community-dwelling adults (n=68) Community setting
Index tests	IQCODE 26 and 17 item, Spanish language
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IIIr
Flow and timing	Of 352 potential subjects, 257 agreed to assessment; 135 completed assessment and data from 68 with suitable informant information were included. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment
Comparative	
Notes	Subjects with moderate to severe dementia were not included, so the study assesses IQCODE against "mild" dementia clinical diagnosis
Methodological quality	

Morales 1995 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
			High
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	Yes		
			Unclear

Morales 1995 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Were missing IQCODE results or un-interpretable IQCODE results reported	Unclear		

Morales 1997 (rural)

Study characteristics			
Patient sampling	Community sampling stratified by age/sex/place of residence with door to door assessment		
Patient characteristics and setting	Community (rural) dwelling adults (n=160)		
Index tests	IQCODE 26 item, Spanish language		
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IIIr		
Flow and timing	Data on numbers assessed and not included are not given. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment		
Comparative			
Notes	This paper presents two separate cohorts; these data refer to those living in a rural setting		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
			High
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	Yes		
			Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Morales 1997 (rural) (*Continued*)

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Were missing IQCODE results or un-interpretable IQCODE results reported	Yes		

Morales 1997 (urban)

Study characteristics			
Patient sampling	Community sampling stratified by age/sex/place of residence with door to door assessment		
Patient characteristics and setting	Community (urban) dwelling adults (n=97) Community setting		
Index tests	IQCODE 26 item, Spanish language		
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IIIr		
Flow and timing	Data on numbers assessed and not included are not given. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment		
Comparative			
Notes	This paper presents two separate cohorts; these data refer to those living in an urban setting		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		

			High
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	Yes		
			Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		

Morales 1997 (urban) (*Continued*)

Were missing IQCODE results or un-interpretable IQCODE results reported	Yes		

Senanorong 2001

Study characteristics			
Patient sampling	"Population" study, no further detail given		
Patient characteristics and setting	Community-dwelling older adults (n=160) Community setting		
Index tests	IQCODE 16 and 3 item, Thai language		
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IV		
Flow and timing	Data on numbers assessed and not included are not given. Timing not applicable as cross-sectional, contemporaneous IQCODE and dementia assessment		
Comparative			
Notes	This study present two cohorts these data are from "normal eduction" group. Numbers of dementia cases suggest a case-control methodology was used but this is not specified in methodology		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			High
DOMAIN 2: Index Test All tests			

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	No		
			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	No		
			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Were missing IQCODE results or un-interpretable IQCODE results reported	Unclear		

Srikanth 2006

Study characteristics			
Patient sampling	Community-based study of all non-aphasic stroke survivors from period 1998-1999 resident in an urban setting		
Patient characteristics and setting	Community-dwelling stroke-survivors (n=79) Community setting		
Index tests	IQCODE 16 item, English language		
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM IV		
Flow and timing	Of 99 subjects, 88 were eligible for assessment and IQCODE data were available for 79		
Comparative			
Notes	These subjects are all stroke-survivors		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			High
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		

Srikanth 2006 (*Continued*)

If a threshold was used, was it pre-specified?	Yes		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	Unclear		
			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	Yes		
			Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Were missing IQCODE results or un-interpretable IQCODE results reported	Yes		

Yamada 2011

Study characteristics			
Patient sampling	Community study, sampling method not clear		
Patient characteristics and setting	Community-dwelling older adults who were participants in another study (n=423)		
Index tests	IQCODE 26 item, Japanese language		
Target condition and reference standard(s)	Clinical dementia diagnosis using DSM		
Flow and timing	Unclear		
Comparative			
Notes	Abstract data only		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
			Unclear
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Were sufficient details given on IQCODE application for the test to be repeated in an independent study	No		

			High
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Were sufficient details of dementia diagnostics given for the assessment to be repeated in an independent sample	No		
			Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Were missing IQCODE results or un-interpretable IQCODE results reported	No		

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

Study	Reason for exclusion
Abreu 2008	Hospital setting
Butt 2008	Data on less than 10 participants

(*Continued*)

Cherbuin 2008	No new data
de Jonge 1997	Data not suitable for analysis
Dekkers 2009	Data not suitable for analysis
Diefeldt 2007b	Repeat data set
Diesfeldt 2007	No dementia diagnosis reference standard
Ehrensperger 2010	Uses unvalidated (two-year) IQCODE
Farias 2002	No dementia diagnosis reference standard
Finneli (abstract)	Data not suitable for analysis
Fuh 1995	Case-control
Garcia 2002	Hospital setting
Goncalves 2011	Hospital setting
Hancock 2009	Hospital setting
Harwood 1997	Hospital setting
Hayden 2003	<10 IQCODE
Henon 2001	Uses a delayed verification analysis
Isella 2002	Uses a delayed verification analysis
Isella 2006	Data not suitable for analysis
Jorm 1989	Data not suitable for analysis
Jorm 1989b	No dementia diagnosis reference standard
Jorm 1991	Hospital setting
Jorm 1996 (Age and Ageing	No dementia diagnosis reference standard
Jorm 1997	No new data
Jorm 2000	No dementia diagnosis reference standard
Jorm 2003	No new data

(*Continued*)

Jorm 2004	No new data
Khachaturian 2000	No IQCODE index test data
Knaefelc 2003	Hospital setting
Krogseth 2011	Uses a delayed verification analysis
Larner 2010	Looks at diagnosis accuracy comparing two dementia types rather than dementia or no dementia dichotomy
Larner 2013	Review article
Li 2012	No dementia diagnosis reference standard
Louis 1999	Uses a delayed verification analysis
Mackinnon 1998	Hospital setting
Mimori (abstract)	No new data
Morales-Gonzalez 1992	Hospital setting
Mulligan 1996	Hospital setting
Narasimhalu 2008	Hospital setting
Ozel-kizel 2010	Hospital setting
Peroco 2009	Hospital setting
Potter 2009	Data not suitable for analysis
Razavi 2011	Hospital setting
Ritchie 1992	No IQCODE data
Rodriguez-Molinero 2010	No dementia diagnosis reference standard
Rovner 2012	Data not suitable for analysis
Sanchez 2009	No dementia diagnosis reference standard
Schofield 2006	Data not suitable for analysis
Sikkes 2010	Hospital setting

(*Continued*)

Siri 2006	Hospital setting
Starr 2000	No dementia diagnosis reference standard
Tang 2003	Hospital setting
Thomas 1994	Hospital setting
Tokuhara 2006	Primary care setting
Wierderholt 1999	Data not suitable for analysis
Wolfe 2009	No dementia diagnosis reference standard
Zevallos-Bustamente 2003	Hospital setting
Zhang 2003	Data not suitable for analysis
Zhou 2002	Hospital setting
Zhou 2003	Repeat data set
Zhou 2004	Repeat data set

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
3 accuracy of IQCODE at 3.3 threshold or nearest (16 and 26 item included)	11	2644
4 accuracy of IQCODE at 3.3 threshold (16 and 26 item IQCODE included)	6	1232
5 accuracy of IQCODE at 3.4 threshold (16 and 26 item IQCODE included)	3	988
6 accuracy of IQCODE at 3.5 threshold (16 and 26 item IQCODE included)	3	1144
7 accuracy of IQCODE at 3.6 threshold (16 and 26 item IQCODE included)	3	1215
8 16 item IQCODE 3.3 threshold	2	763
9 16 item IQCODE 3.4 threshold	3	988
10 16 item IQCODE 3.5 threshold	1	684
11 16 item IQCODE 3.6 threshold	1	646
12 26 item IQCODE 3.3 threshold	5	1153
13 26 item IQCODE 3.4 threshold	1	674
14 26 item IQCODE 3.5 threshold	2	460
15 26 item IQCODE 3.6 threshold	2	569
16 all IQCODE studies at 3. 3 threshold with Srikanth removed	5	1153
17 all IQCODE studies at 3.4 threshold with Senanorong removed	2	828
18 all IQCODE studies at 3.5 threshold with Senanorong removed	3	1144

ADDITIONAL TABLES

Table 1. Summary of test accuracy at study level

Study ID	Participants (n)	Primary threshold	Sensitivity (%)	Specificity (%)
Jorm 1994	684	3.4	77	86
Jorm (psychiatry) 1996	144	3.3	91	62
Kathriarachi 2001	37	3.5	71	83
Law 1995	237	3.6	75	98
Mackinnon 2003	646	3.6	67	93
Morales 1995	68	3.3	86	92
Morales (rural) 1997	160	3.3	82	90
Morales (urban) 1997	97	3.3	83	83
Senanorong 2001	160	3.5	85	92
Srikanth 2006	79	3.3	88	63
Yamada 2011	423	3.6	80	85

Table 2. Age of participants in included studies

Study name	Mean age (yrs)	SD	Range (yrs)
Jorm 1994	-	-	-
Jorm 1996 (psychiatry)	72.9	-	66 - 83
Kathriarachi 2001	-	-	*(Recruited >65yrs only)*
Law 1995	80.7	6.5	67 - 97
Mackinnon 2003	76.5	-	70 - 97
Morales 1995	73.1	5.2	65 - 86
Morales 1997 (urban) (urban)	75.2	6.1	66 - 92
Morales 1997 (urban) (rural)	73.5	8.2	61 - 96
Senanorong 2001	65.7	5.0	52 - 85

Table 2. Age of participants in included studies (*Continued*)

Srikanth 2006	69.0	14.4	-
Yamada 2011	-	-	-

CONTRIBUTIONS OF AUTHORS

ANS assisted with search terms and translation; TJQ and PF drafted the protocol, performed data extraction and quality assessment; CY assisted with data handling; RMcS and DJS provided support and contributed to manuscript content.

DECLARATIONS OF INTEREST

The review authors have no declarations specific to the review.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- RCPSG Travelling Fellowship, UK.

Dr Quinn spent a period working directly with the Cochrane Group; this was supported by a RCPSG travelling fellowship.

- Graham Wilson Travelling Scholarship, UK.

Dr Quinn's travel and accomodation to allow training in review methodology and analysis was supported by a Graham Wilson Travelling Scholarship.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

All differences between the protocol and review are described in the main body of the text. A priori we had planned a number of covariate and sensitivity analyses, however the data set was limited in numbers of studies and studies were too heterogenous to allow all of our planned analyses. We had originally planned to review the test accuracy of the 16 and 26-item IQCODE separately, however given the modest number of studies and a comparative analysis suggesting no systematic difference between the two IQCODE formats we used pooled data for our primary analysis.