Creavin, S. T., Noel-Storr, A. H., Richard, E., Creavin, A. L., Cullum, S., Ben-Shlomo, Y., & Purdy, S. (2017). Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people. *Cochrane Database of Systematic Reviews*, *2017*(2), [CD012558]. https://doi.org/10.1002/14651858.CD012558

Cochrane Library

Cochrane Database of Systematic Reviews

Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people (Protocol)

Creavin ST, Noel-Storr AH, Richard E, Creavin AL, Cullum S, Ben-Shlomo Y, Purdy S

Creavin ST, Noel-Storr AH, Richard E, Creavin AL, Cullum S, Ben-Shlomo Y, Purdy S.
Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people.
Cochrane Database of Systematic Reviews 2017, Issue 2. Art. No.: CD012558.
DOI: 10.1002/14651858.CD012558.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people

Sam T Creavin[1], Anna H Noel-Storr[2], Edo Richard[3], Alexandra L Creavin[1], Sarah Cullum[4], Yoav Ben-Shlomo[5], Sarah Purdy[6]

[1]School of Social and Community Medicine, University of Bristol, Bristol, UK. [2]Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [3]Department of Neurology, Radboud University Nijmegen Medical Center, Nijmegen, Netherlands. [4]Department of Psychological Medicine, University of Auckland, Auckland, New Zealand. [5]Dept of Social Medicine, Canynge Hall, Bristol, UK. [6]Faculty of Health Sciences, University of Bristol, Bristol, UK

Contact address: Sam T Creavin, School of Social and Community Medicine, University of Bristol, Carynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK. sam.creavin@bristol.ac.uk.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.
Publication status and date: New, published in Issue 2, 2017.

Citation: Creavin ST, Noel-Storr AH, Richard E, Creavin AL, Cullum S, Ben-Shlomo Y, Purdy S. Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people. *Cochrane Database of Systematic Reviews* 2017, Issue 2. Art. No.: CD012558. DOI: 10.1002/14651858.CD012558.

ABSTRACT

This is a protocol for a Cochrane Review (Diagnostic test accuracy). The objectives are as follows:

To determine the accuracy of general practitioners' overall gestalt (unaided) clinical judgement for diagnosing cognitive impairment and dementia in symptomatic people presenting to primary care. There is no comparator index test.

To investigate the heterogeneity of test accuracy in the included studies.

BACKGROUND

Cochrane is undertaking a series of reviews investigating the diagnostic accuracy of a variety of tests for diagnosing dementia, but to contextualise the findings to practice it is also important to quantify the accuracy of clinical judgement. Doctors use a variety of processes to reach a diagnosis, including non-analytical reasoning processes such as pattern recognition, to rapidly generate diagnostic hypotheses (Norman 2007; Elstein 2009). Some people with dementia unfortunately have sufficiently advanced disease at the point of diagnosis that additional tests may be unnecessary and burdensome. General practitioners (GPs) often report using their clinical judgement, rather than a formal test, to determine whether someone has dementia (O'Connor 1993; Pentzek 2009). A review of the clinical judgement of GPs is therefore an important step in determining the potential added value of more formal diagnostic workup, such as brief cognitive tests.

Target condition being diagnosed

In this protocol we investigate the accuracy of gestalt clinical judgement for the diagnosis of two target conditions: all-cause dementia, and cognitive impairment due to dementia or mild cognitive impairment (MCI).

Dementia is a clinical syndrome of cognitive impairment that develops gradually and causes a decline in functioning. Dementia is increasingly common with age, affecting less than 5% of the population aged less than 75 years and 17% of those aged over 89 years (Matthews 2013). Dementia may result from a variety of pathologies, but in the elderly population in the community these subtype definitions based on disease aetiology are thought by some investigators to be of less relevance, as most old people with dementia have mixed pathology at autopsy (Neuropathology 2001; Savva 2009; Brayne 2012; Kawas 2015).

Cognitive impairment includes dementia and MCI (Gauthier 2006). MCI is a syndrome of cognitive impairment that is greater than expected when accounting for a person's age and educational attainment, but that does not interfere with capacity for independence in everyday activities of daily living. MCI affects between 3% and 20% of adults aged over 65 years (Gauthier 2006), and the prognosis in general practice is variable: approximately 25% of people develop dementia within three years but around 40% revert to normal (Kaduszkiewicz 2014).

Experience in clinical general practice is that when there are concerns about impaired cognition these are focused primarily on the possibility of dementia rather than MCI, but inevitably some people who are evaluated for possible dementia will be diagnosed with MCI. In this protocol we include people who are ultimately diagnosed as having MCI when we refer to a person consulting with a GP about possible dementia (e.g. under Participants or Clinical pathway), because it would be unusual for a person to consult a GP about possible MCI or cognitive impairment. Our second target condition includes both dementia *and* MCI because it would be unusual for a GP to diagnose MCI, especially on the basis of gestalt judgement alone, because neuropsychological evaluation is often required. If gestalt clinical judgement was sensitive for any cognitive impairment, then if the GP assessed the person as being cognitively normal it would rule out both dementia and MCI.

Index test(s)

The index test will be a clinical diagnosis of cognitive impairment (due to MCI or dementia), or dementia, based on the overall clinical judgement (or gut feeling/gestalt (Lehman 2015)) of a primary care physician after a clinical assessment, unaided by formal (even brief) cognitive tests. We operationalise this as a single index test (clinical judgement) with two target conditions (see Target condition being diagnosed for details). Diagnostic labels in general practice may function primarily to guide the management of the patient, to treat, to investigate, or to exclude serious disease (Jones 2010). GPs have been described as using intuition (Barraclough 2006; Woolley 2013), pattern recognition (Heneghan 2009) and scripts (Charlin 2000), amongst other strategies (Heneghan 2009), to reach a diagnosis.

The diagnostic accuracy of GPs' clinical judgment about the presence of dementia after consulting with patients had good diagnostic accuracy (sensitivity 92%, specificity 76%) in one study (Cooper 1992). This compared fairly well to the diagnostic accuracy of the informant questionnaire for cognitive disorders in the elderly (IQCODE), a brief cognitive test for diagnosing dementia, at a cutpoint of 3.2 (sensitivity 100%, specificity 76%) (Harrison 2014) and, in a different clinical context, to the clinical judgement of GPs regarding the severity of chest pain aetiology based only on brief history and examination (sensitivity 82%, specificity 79%) (Buntinx 1991). GPs report lack of time as a barrier to diagnosing dementia (Koch 2010) and report often relying on personal observations to make the diagnosis (O'Connor 1993) whereas pen-and-paper tests are used by a minority of people (Pond 2013).

Clinical pathway

Prior tests

Many people who are concerned about the onset of possible dementia present to a healthcare provider for an evaluation; often the first consultation would be with a primary care provider (commonly a GP) but in some health economies the first consultation may be with a specialist clinician. Some people may not experience subjective cognitive problems (Waldorff 2012), but may be encouraged (or taken) to attend a consultation with a clinician by a close contact (e.g. a carer) or professional who is concerned about possible dementia. A further possibility is that a GP may form an impression of possible cognitive impairment during a consultation with a patient about a (potentially) unrelated matter.

Most commonly in research studies and clinical practice, no tests would be performed before a GP consultation regarding possible dementia. Some people may consult with their GP about the possibility of dementia after performing a self-administered cognitive test such as test-your-memory (Brown 2009). Alternatively some people might have been asked to see their GP as a consequence of undergoing brief cognitive testing conducted by another health professional (for example, a district nurse or hospital doctor), or as part of a research project.

In this review we will only consider clinical judgement by a primary care physician (GP) in someone who is considered to have symptoms. Either the patient themselves or someone else, including a health professional (including the consulting GP), should be concerned about possible cognitive impairment. Recent policy in the USA and UK has encouraged screening for dementia in people who do not have symptoms (Burns 2013; Rasmussen 2013; Rasmussen 2014). This remains controversial (Brayne 2007; Fox 2013; Le 2013; Iliffe 2014) and we do not propose to include these people in this review.

Role of index tests

It is rare that any single component of a diagnostic evaluation would be diagnostic for a condition by itself. Most people who are being evaluated for possible cognitive impairment or dementia will commonly undergo further assessment that may include brief cognitive tests and investigations such as biochemical analysis and neuroimaging. Most commonly in primary care, further assessment of a patient is dependent on the GPs' clinical judgment: when the GP feels comfortable to exclude cognitive impairment or dementia without further assessment the patient will usually undergo no further tests, whereas when the GP feels uncertain then further evaluation may be arranged. It would be unusual for a GP to rule-in dementia without further assessment, but this may occur when the patient is frail, affected by multiple comorbidities, and perhaps resident in a nursing home, where the prior probability of dementia (or prevalence) may be as high as 60% (Magaziner 2000), and when the management may be primarily palliative. In some situations GPs may use a test of time (Almond 2009) to help increase the specificity of a diagnosis, especially when the condition may fluctuate, and a GP may therefore form a 'working diagnosis', which is reviewed over a period of time, before deciding on a formal recorded diagnosis.

Alternative test(s)

Alternatives to the index test would include a more detailed evaluation, which may be conducted by a specialist, and might include aspects of clinical history, examination, cognitive testing, biochemical and haematological analysis and neuroimaging.

Rationale

A systematic review published in 2010 found that the judgement of GPs was highly specific for diagnosing dementia at all stages of severity, but only moderately sensitive (van den Dungen 2012). A second review addressing a similar question used a more restricted search strategy (Mitchell 2011). Both reviews were well conducted but allowed a broad definition of 'clinical judgement' that is not immediately applicable to clinical practice, by including studies where 'clinical judgement' was defined as a documented diagnosis in the medical records, which may not accurately reflect the actual clinical opinion (Russell 2013). Additionally, there is scope to develop the search strategy, in particular to include more terms relating to dementia, cognitive impairment and diagnostic accuracy.

OBJECTIVES

To determine the accuracy of general practitioners' overall gestalt (unaided) clinical judgement for diagnosing cognitive impairment and dementia in symptomatic people presenting to primary care. There is no comparator index test.

Secondary objectives

To investigate the heterogeneity of test accuracy in the included studies.

METHODS

Criteria for considering studies for this review

Types of studies

We will include cross-sectional studies (where participants have index test and reference test at the same encounter, which would be unusual) and cohort studies. We recognise that cross-sectional studies might be at higher risk of incorporation bias than cohort studies and we will account for this when we assess studies for risk of bias (quality appraisal); we judge that the alternative approach of excluding cross-sectional studies would be too restrictive. We will not include case-control studies because they are at high risk of bias and because, by definition, any participants would have been recruited on the basis of disease state (dementia, cognitive impairment or normal). This would prevent GPs from making a blinded gestalt clinical judgement about the diagnosis, because in most health systems the GP primary care record contains entries relating to all medical and psychiatric diagnoses, which would include cognitive impairment and dementia.

Participants

We will only include studies that have recruited participants from primary care. We define primary care as first-contact health care provided by a non-specialist clinician in a continuing-care office setting. We will exclude studies where the consultation with a non-specialist takes place in hospital (including outpatients or emergency departments) as this is unlikely to represent primary care in the sense that is relevant to our review. Because we anticipate that reporting in original studies may be suboptimal (Noel-Storr 2014), we will include studies where some or all of the participants are consulting with a primary care provider about possible dementia following a recommendation by a non-specialist secondary care provider (e.g. emergency department), even if the study does not explicitly state these people were consulting secondary care about a non-dementia concern (e.g. a fall).

We will only include studies where GPs make a clinical judgement about the presence or absence of cognitive impairment or dementia in someone who is suspected of having it (either by the patient,

a caregiver, or professional - including the consulting GP). We will exclude studies where GPs are asked to make a judgement about the presence or absence of cognitive impairment or dementia in all people attending primary care, regardless of the reason for attendance, as this is akin to screening. We recognise that other primary care providers might form clinical judgements about the possibility of cognitive impairment, but we will only include studies that use GPs as the primary care provider because we are concerned that including other professionals would introduce even greater heterogeneity than we already anticipate; for instance, although GPs are required to hold a license to practice medicine, the training requirements and scope of practice may vary substantially in different countries.

Index tests

We propose that the core feature of clinical judgement is that it is unaided by any additional test, investigation or inquiry beyond that which is immediately available to the clinician (Blaeuer 2013; Di 2013; Body 2014). As outlined above, in this review we are investigating a single index test (clinical judgement) with two target conditions (cognitive impairment composite or dementia). In everyday practice, a clinical judgement is necessarily formed after an encounter with a patient during which GPs would often have access to the medical record and might review this in conjunction with meeting the patient. There are three ways that clinical judgement (for research) may be used in a diagnostic accuracy study in general practice. The first definition is a documented diagnosis of cognitive impairment or dementia in the medical records; we consider that this definition reflects the process of documentation rather than clinical judgement. The second definition is a judgement of a clinician based on knowledge of the patient and review of the medical notes, but not relating to a specific encounter with the patient; we consider that this definition reflects consulting behaviour of people (in this case with cognitive impairment or dementia). The third definition is a clinical impression formed by the clinician after consulting with a patient who has presented to a specific encounter with the doctor (perhaps with symptoms suggestive of possible dementia, though not always - because it may be the consulting GP who raises the possibility of cognitive problems), and we consider this to be the definition of clinical judgement that is most relevant to practice. For this review, we will include studies that use the third definition (clinical impression after consultation) but we anticipate that there will be very few studies that use this design. Therefore, to avoid an empty review, we will also include studies that use the second definition (based on existing knowledge of the patient and not relating to a particular encounter) so long as the index test (GP judgement about cognitive impairment or dementia) has taken place before any definitive diagnosis (for example, specialist assessment in a memory clinic). We will investigate the use of medical records as a source of heterogeneity under the category 'prior tests' for any study that allows doctors access to the medical records, regardless of whether clinical judgement is defined using definition two or three. For studies that use definition two it will usually be explicit that GPs were allowed to review the medical record, but if this is not clear we will always make the assumption that the records were reviewed. We will judge that access to the medical records was allowed for studies that use definition three only if this is explicit. The doctor's clinical impression will often determine the extent of the additional work-up offered. In one scenario, people who are thought to be highly likely to have dementia might have only a brief 'rule-in' test together with blood tests to exclude other causes such as hypothyroidism or infection, or (rarely) no additional tests at all; this scenario is less applicable to people who are thought to have cognitive impairment rather than frank dementia. In a second scenario, where there is a degree of uncertainty, people might be referred to a specialist, and in a third scenario those who are thought to be highly unlikely to have any problems might be offered a brief 'rule-out' test, or none at all. We will include studies where some (but not all) participants undergo both the index test and reference standard, so long as at least some index test positives and index test negatives undergo the reference standard, and will account for this verification bias using the QUADAS-2 checklist; for these studies we will use the population undergoing both tests as the denominator for diagnostic accuracy and we will document the prevalence of cognitive impairment and dementia in the total sample separately. We will not exclude studies where GPs are allowed to use additional cognitive tests to help determine the management of the patient *after* formulating and expressing their unaided judgement, but if we judge that these additional tests have contributed to formulation of clinical judgement we will account for this as a source of heterogeneity as described below. However, we will not evaluate the accuracy of any tests other than clinical judgement in this review.

Original studies may offer GPs two (cognitive impairment, normal; or dementia, normal) or three possible diagnostic categories (dementia, cognitive impairment, normal), and may ask GPs to rate their confidence in the diagnosis, or how probable it is.

Impact of GP decision-making on further evaluation and verification bias

If the GP judges that cognitive impairment or dementia is unlikely and does not perform any further verification, but the person has cognitive impairment or dementia, then this person would be a false negative case; in this event prevalence of cognitive problems will be underestimated, and estimated sensitivity will be higher than the true value. In other cases GPs might suspect dementia but not take any further action to confirm (with further tests) or to document the diagnosis; in this event prevalence will be underestimated and the estimated sensitivity and specificity would both be affected (most likely they will be underestimated but this is impossible to determine). This second circumstance might occur

if the doctor is not specifically asked about their judgement about dementia after consulting with patients, because in general practice the specific diagnosis is often less important than the prognosis and impact for the patient (Schellevis 2004). This is more likely if the patient is predominantly burdened by some other physical health problem and has an anticipated short life expectancy, so that the purpose of confirming a dementia diagnosis may be unclear (Slavin 2013). This situation is likely to be much more of an issue for people with dementia than those with cognitive impairment. Only research studies that offer a reference standard assessment to all people presenting with concerns about cognitive impairment, regardless of the GPs' gestalt clinical judgement, will be able to report robust data on diagnostic accuracy.

Target conditions

The first target condition is all-cause dementia. We will include a diagnosis of dementia at any stage of disease, because we do not want to restrict our results and this pragmatic approach is most relevant to clinical practice. We will not examine the utility of clinical judgement for risk prediction of future dementia. The second target condition is cognitive impairment due to MCI or dementia.

Reference standards

To allow for a pragmatic and sensitive approach to study inclusion, we will include different reference standards (outlined below). Studies must administer the index test and reference test (excepting longitudinal follow up) within six months; if authors do not provide details of this time interval we will include the study and account for this as 'unclear' in the quality appraisal using QUADAS-2.

Dementia

We will include studies that apply the reference standard of all-cause dementia according to DSM (American Psychiatric Association) or ICD (ICD 1993) definitions, regardless of version. We will also include studies that use Agecat (Copeland 1986), CAMDEX (Roth 1986) and Clinical Dementia Rating Scale (Hughes 1982) as the reference standard, as these are well-validated methods of applying the aforementioned diagnostic criteria. We will include studies that use expert specialist clinical judgement as the reference standard. We consider a specialist to be a clinician who has particular expertise in diagnosing and managing dementia, who will usually practice in a hospital, and have the professional status of a geriatrician, psychiatrist or neurologist. We will include studies that use longitudinal confirmation of the diagnosis of all-cause dementia in primary care, because we anticipate that in some studies a specialist assessment will only be offered to some participants. We operationalise 'longitudinal confirmation of the diagnosis in primary care' as case record review occurring at least three months *after* the index test diagnosis of dementia where no other alternative diagnosis is identified. It is likely that many people who can be correctly diagnosed as having dementia by unaided clinical judgement (true positives) would have a fairly advanced stage of disease, but stage of disease will not form part of the target condition.

Although the target condition is all-cause dementia we will also include studies that use an aetiological sub-type definition: for Alzheimer disease dementia (McKhann 1984; McKhann 2011), vascular dementia (Román 1993), Lewy body dementia (McKeith 1996; McKeith 2005) or frontotemporal dementia (Neary 1994).

Cognitive impairment

Cognitive impairment is a composite target condition. We will allow any recognised definition of MCI (Petersen 1999; Petersen 2004; Winblad 2004; McKhann 2011), as well as the reference standards for dementia outlined above.

In addition to dementia and MCI there are other causes of cognitive impairment, such as delirium and head injury, but these are not part of the target condition that we are investigating in this review. If the index tests indicated cognitive impairment or dementia and further evaluation demonstrated that the clinical problem was delirium instead, the test would be false positive.

Search methods for identification of studies

Electronic searches

We will search MEDLINE (OvidSP); Embase (OvidSP); BIOSIS previews (Thomson Reuters Web of Science); Web of Science Core Collection, including the Science Citation Index and the Conference Proceedings Citation Index (Thomson Reuters Web of Science); PsycINFO (OvidSP), LILACS (BIREME) and ALOIS (www.medicine.ox.ac.uk/alois). See Appendix 1 for the MEDLINE search strategy. Where appropriate, we will use controlled vocabulary such as MeSH terms (in MEDLINE) and EMTREE (in Embase) and other controlled vocabulary in other databases, as appropriate.

Search filters are collections of terms aimed at reducing the number needed to screen by filtering out irrelevant records and retaining only those that are relevant. We will not use search filters designed to retrieve diagnostic test accuracy studies as a method to restrict the search overall, because available filters have not yet proved sensitive enough for systematic review searches (Whiting 2011a). We will include a validated filter for primary care studies that optimises sensitivity and specificity (Gill 2014). We will not apply any language restriction to the electronic searches.

Searching other resources

We will check the reference lists of all relevant papers for additional studies. We will also search:

- meta-analyses van Diagnostisch Onderzoek (MEDION database) (www.mediondatabase.nl);
- NIHR Dissemination Centre (which replaced DARE) (discover.dc.nihr.ac.uk/portal);
- Health Technology Assessment Database (HTA Database) in the Cochrane Library (www.cochranelibrary.com); and
- Aggressive Research Intelligence Facility (ARIF database) (147.188.28.230/rmwp).

We will also talk to experts and attempt to contact authors where necessary to obtain details of unpublished studies.

Data collection and analysis

Selection of studies

One review author will screen all retrieved titles for relevance and classify titles as definitely relevant, possibly relevant and definitely irrelevant; possibly relevant titles will be considered by a second author to determine whether the abstract should be reviewed (default position) or not. Two authors will then assess all relevant abstracts, resolving disagreements about whether to include an article by discussion and by involving an arbiter where necessary. We will attempt to retrieve any potentially eligible studies for full text review.

If data from a study are presented in multiple papers we will present this under a 'primary reference' based on the study that provided most data to our review, unless papers contribute similar amounts of data, in which case we will designate the primary reference based on publication date of manuscripts. We will detail study selection in a PRISMA flowchart. We will attempt to categorise reasons for excluding articles at the full text stage under the following hierarchy.

1. **Inappropriate participants**
   i) Not primary care
   ii) Index test not performed in someone where there is a suspicion of dementia (i.e. a screening study)
2. **Inappropriate reference standard**
   i) Not one of the specified reference standards
3. **Inappropriate index test**
   i) Not GP
   ii) Not gestalt clinical judgement
4. **Inappropriate target condition**
5. **Inappropriate study design** (i.e. not a diagnostic test accuracy study e.g. a study reporting qualitative data, descriptive epidemiology, randomised trial or survey)

Difficulties can arise in reviews of diagnostic accuracy as to whether to include studies where information on diagnostic accuracy on the index test of interest *might* be available but is not reported. Table 1 shows the circumstances under which we will contact authors in the hope of obtaining relevant information on diagnostic accuracy.

Data extraction and management

We will use a study specific pro-forma to extract information based on the list required for Cochrane reviews of diagnostic test accuracy: sampling, characteristics of participants and setting, index test, target condition, reference test, flow and timing, use of prior tests and comparator tests. We will also extract data relating to study level covariates of average age, proportion of women participants, average scores on any cognitive test, stage or severity of dementia, average educational attainment for participants, average age and experience of general practitioners performing index test, and proportion of male and female doctors. We will also extract study level covariates relating to country of study and type of practice (categorised as single, group, teaching/academic).

We will extract information relating to the index test based on what is available in the primary study, which may include both or either target conditions. There is no accepted cut point for the index test so we will use the binary classification of whether the GP judges dementia to be present (index test positive for target condition dementia) or not (index test negative for target condition dementia), and similarly for cognitive impairment as the target condition. Where the judgement of the GP is expressed as a probability we will consider probabilities of 51% and more as indicating the target condition is considered present (index test positive). We will extract all the relevant data including, where reported, results for both all-cause dementia and aetiological subtypes.

We will contact authors of included primary studies to obtain missing or unclear information relating to covariates listed above and/or items on the QUADAS-2 checklist.

Assessment of methodological quality

Two authors will assess study quality using the QUADAS-2 checklist (Whiting 2011) separately and disagreements will be resolved by discussion and involvement of an arbitrator if necessary.

Statistical analysis and data synthesis

We will use paired data on sensitivity and specificity to calculate the accuracy of the index test for diagnosing the two target conditions: cognitive impairment (including both MCI and all-cause dementia), and all-cause dementia. We will calculate the diagnostic accuracy with 95% confidence intervals separately for each target condition in all studies with available data.

We will perform meta-analyses on pairs of sensitivity and specificity, if it is appropriate to pool the data, using the bivariate random-effects model approach based on pairs of sensitivity and specificity (Reitsma 2005; Chu 2006; Harbord 2007; Macaskill

2010). We will use Stata software (StataCorp 2013) to carry out the additional analyses using the bivariate approach. If it is not appropriate to perform meta-analysis we will synthesise the results narratively (Ryan 2013). We will only perform meta-analyses using all-cause dementia diagnostic criteria (DSM, ICD, Agecat, CAMDEX, CDR) as this is the target condition that is most applicable to primary care (rather than aetiological subtypes), and because in elderly patients there is often mixed pathology. We will combine different all-cause dementia diagnostic criteria for the five listed all-cause dementia diagnostic criteria. Expert diagnosis of all-cause dementia that does not meet one of the listed research definitions will be meta-analysed separately if appropriate. We will not perform meta-analyses by aetiological subtype of dementia.

We will also perform meta-analysis with the composite target outcome of cognitive impairment (including MCI and all-cause dementia). In this analysis true positives will be all cases who are identified by one of our applicable Reference standards as having either MCI or all-cause dementia.

If more than one study reports data for the index test (judgement of GPs) as a probability then we will model this as an implicit threshold in meta analyses.

Investigations of heterogeneity

We will investigate two sources of heterogeneity: the use of prior tests or medical records, and the number of diagnostic categories that are available to GPs in the original study. We consider that medical records can be conceptualised as a prior test, and that diagnostic accuracy might be influenced by whether an original study offers GPs three possible diagnostic categories (dementia, cognitive impairment, normal) rather than two (cognitive impairment, normal; or dementia, normal). We will initially investigate heterogeneity through visual examination of forest plots - of sensitivities and specificities - and the ROC plot of the raw data. Where there is evidence of heterogeneity we will attempt to adjust for this in the model through inclusion in the hierarchical regression model. We will use likelihood ratio tests to compare model fit.

We will specifically not include the length of training or type of training programme as sources of heterogeneity, as we anticipate these will be poorly reported in original studies and hard to obtain information if we contact authors. We will not adjust for study characteristics that are only reported as aggregate measures (e.g. mean scores of cognitive testing), as it is recommended to only investigate heterogeneity in diagnostic accuracy by characteristics that can be assessed at the study level (Bossuyt 2013).

Sensitivity analyses

We will investigate how our estimates of diagnostic accuracy are modified when we exclude studies that are judged to be at high risk of bias in more than two domains, or that use extended primary care follow up or expert clinical judgement as the reference standard, from the analysis.

Assessment of reporting bias

Quantitative methods for exploring reporting bias are not well established for studies of DTA (Bossuyt 2013) and so we will not investigate reporting bias.

ACKNOWLEDGEMENTS

None

REFERENCES

Additional references

Almond 2009
Almond SC, Summerton N. Diagnosis in general practice. Test of time. *BMJ* 2009;**15**(338):b1878. [PUBMED: 19528115]

American Psychiatric Association
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders Fourth TR*. Arlington (VA): American Psychiatric Association, 2000.

Barraclough 2006
Barraclough K. Medical intuition. *BMJ* 2006;**332**:497.

Blaeuer 2013
Blaeuer SR, Bally K, Tschudi P, Martina B, Zeller A. Acute cough illness in general practice - predictive value of clinical judgement and accuracy of requesting chest x-rays. *Praxis* 2013;**102**(21):1287–92.

Body 2014
Body R, Cook G, Burrows G, Carley S, Lewis PS. Can emergency physicians "rule in" and "rule out" acute myocardial infarction with clinical judgement?. *Emergency Medicine Journal* 2014;**31**(11):872–6.

Bossuyt 2013
Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Chapter 11: Interpreting results and drawing conclusions. In: Deeks JJ, Bossuyt PM, Gatsonis C editor (s). *Cochrane Handbook for Systematic Reviews of Diagnostic test Accuracy Version 0.9*. The Cochrane Collaboration, 2013:1–31.

Brayne 2007
Brayne C, Fox C, Boustani M. Dementia screening in primary care: is it time?. *JAMA* 2007;**298**(20):2409–11.

Brayne 2012
Brayne C, Davis D. Making Alzheimer's and dementia research fit for populations. *Lancet* 2012;**380**(9851): 1441–3.

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

Buntinx 1991
Buntinx F, Truyen J, Embrechts P, Moreel G, Peeters R. Chest pain: an evaluation of the initial diagnosis made by 25 Flemish general practitioners. *Family Practice* 1991;**8**(2): 121–4.

Burns 2013
Burns A. Alistair Burns and 51 colleagues reply to David Le Couteur and colleagues. *BMJ (Clinical research ed.)* 2013; **347**(oct15_6):f6125.

Charlin 2000
Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Academic Medicine* 2000;**75**(1):182–90.

Chu 2006
Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006;**59** (12):1331-2; author reply 1332-3.

Cooper 1992
Cooper B, Bickel H, Schäufele M. The ability of general practitioners to detect dementia and cognitive impairment in their elderly patients: a study in Mannheim. *International Journal of Geriatric Psychiatry* 1992;**7**(8):591–8.

Copeland 1986
Copeland JR, Dewey ME, Griffiths-Jones HM. A computerized psychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGECAT. *Psychological Medicine* 1986;**16**(1):89–99.

Di 2013
Di Somma S, Magrini L, De Berardinis B, Marino R, Ferri E, Moscatelli P, et al. Additive value of blood neutrophil gelatinase-associated lipocalin to clinical judgement in acute kidney injury diagnosis and mortality prediction in patients hospitalized from the emergency department. *Critical Care* 2013;**17**(1):R29.

Elstein 2009
Elstein AS. Thinking about diagnostic thinking: a 30-year perspective. *Advances in Health Sciences Education* 2009;**14** (1 Suppl):7–18.

Fox 2013
Fox C, Lafortune L, Boustani M, Brayne C. The pros and cons of early diagnosis in dementia. *British Journal of General Practice* 2013;**63**(612):e510–2.

Gauthier 2006
Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. *Lancet* 2006; **367**(9518):1262–70.

Gill 2014
Gill PJ, Roberts NW, Wang KY, Heneghan C. Development of a search filter for identifying studies completed in primary care. *Family Practice* 2014;**31**(6):739–45.

Harbord 2007
Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;**8**(2):239–51.

Harrison 2014
Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting. *Cochrane Database of Systematic Reviews* 2014, Issue 7. [DOI: 10.1002/14651858.CD010771.pub2]

Heneghan 2009
Heneghan C, Glasziou P, Thompson M, Rose P, Balla J, Lasserson D, et al. Diagnostic strategies used in primary care. *BMJ (Clinical research ed.)* 2009;**338**(April):b946.

Hughes 1982
Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *British Journal of Psychiatry* 1982;**140**:566–72.

ICD 1993
World Health Organization. *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. Geneva: World Health Organization, 1993.

Iliffe 2014
Iliffe S. General practitioners should be conducting targeted screening for dementia in people aged 65 to 74. *Journal of Primary Health Care* 2014;**6**(3):247–9.

Jones 2010
Jones R, Barraclough K, Dowrick C. When no diagnostic label is applied. *BMJ (Clinical research ed.)* 2010;**340**(June): c2683.

Kaduszkiewicz 2014
Kaduszkiewicz H, Eisele M, Wiese B, Prokein J, Luppa M, Luck T, et al. Prognosis of mild cognitive impairment in general practice: results of the German AgeCoDe study. *Annals of Family Medicine* 2014;**12**(2):158–65.

Kawas 2015
Kawas CH, Kim RC, Sonnen J, Bullain SS, Trieu T. Multiple pathologies are common and related to dementia in the oldest-old. *Neurology* 2015;**85**(6):535–42.

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

Le 2013
Le Couteur DG, Doust J, Creasey H, Brayne C. Political drive to screen for pre-dementia: not evidence based and

ignores the harms of diagnosis. *BMJ (Clinical research ed.)* 2013;**347**:f5125.

Lehman 2015
Lehman R. Siddharta Mukherjee's three laws of medicine. *BMJ* 2015;**351:h6708**:1–2.

Macaskill 2010
Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C editor(s). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Cochrane Collaboration, 2010.

Magaziner 2000
Magaziner J, German P, Zimmerman SI, Hebel J R, Burton L, Gruber-Baldini AL, et al. The prevalence of dementia in a statewide sample of new nursing home admissions aged 65 and older: diagnosis by expert panel. Epidemiology of Dementia in Nursing Homes Research Group. *Gerontologist* 2000;**40**(6):663–72.

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

McKeith 1996
McKeith IG, Galasko D, Kosaka K, Perry EK, Dickson D W, Hansen LA, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology* 1996;**47**(5):1113–24.

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

McKhann 1984
McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;**34**(7): 939–44.

McKhann 2011
McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* 2011;**7**(3):263–9.

Mitchell 2011
Mitchell AJ, Meader N, Pentzek M. Clinical recognition of dementia and cognitive impairment in primary care: a meta-analysis of physician accuracy. *Acta Psychiatrica Scandinavica* 2011;**124**(3):165–83.

Neary 1994
Neary D, Brun A, Englund B, Gustafson L, Passant U, Mann DMA, et al. Clinical and neuropathological criteria for frontotemporal dementia. The Lund and Manchester Groups. *Journal of Neurology, Neurosurgery, and Psychiatry* 1994;**57**(4):416–8.

Neuropathology 2001
Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. *Lancet* 2001;**357**(9251):169–75.

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

Norman 2007
Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Medical Education* 2007;**41**(12):1140–5.

O'Connor 1993
O'Connor DW, Fertig A, Grande MJ, Hyde JB, Perry JR, Roland MO, et al. Dementia in general practice: the practical consequences of a more positive approach to diagnosis. *Journal of the Royal College of General Practitioners* 1993;**43**(370):185–8.

Pentzek 2009
Pentzek M, Fuchs A, Wiese B, Cvetanovska-Pllashniku G, Haller F, Maier W, et al. AgeCoDe study group. General practitioners' judgment of their elderly patients' cognitive status. *Journal of General Internal Medicine* 2009;**24**(12): 1314–7. [PUBMED: 19844763]

Petersen 1999
Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* 1999; **56**(3):303–8.

Petersen 2004
Petersen RC. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine* 2004;**256 (3)**:183–94.

Pond 2013
Pond CD, Mate KE, Phillips J, Stocks NP, Magin PJ, Weaver N, et al. Predictors of agreement between general practitioner detection of dementia and the revised Cambridge Cognitive Assessment (CAMCOG-R). *International Psychogeriatrics* 2013;**25**(10):1639–47.

Rasmussen 2013
Rasmussen J. Improving diagnosis and management of dementia in primary care. *Progress in Neurology and Psychiatry* 2013;**17**(6):4–6.

Rasmussen 2014
Rasmussen J. General practitioners should be conducting targeted screening for dementia in people aged 65 to 74: yes. *Journal of Primary Health Care* 2014;**6**(3):245–7.

Reitsma 2005
Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–90.

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

Roth 1986
Roth M, Tym E, Mountjoy CQ, Huppert FA, Hendrie H, Verma S, et al. CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *British Journal of Psychiatry* 1986;**149**:698–709.

Russell 2013
Russell P, Banerjee S, Watt J, Adleman R, Agoe B, Burnie N, et al. Improving the identification of people with dementia in primary care: evaluation of the impact of primary care dementia coding guidance on identified prevalence. *BMJ Open* 2013;**3**(12):e004023.

Ryan 2013
Ryan R, Cochrane Consumers and Communication Review Group. [Cochrane Consumers and Communication Review Group: data synthesis and analysis]. cccrg.cochrane.org (accessed 14 February 2017) June 2013.

Savva 2009
Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C. Age, neuropathology, and dementia. *New England Journal of Medicine* 2009;**360**(22):2302–9.

Schellevis 2004
Schellevis FG. Physical and mental illness. In: Jones R, Britten N, Culpepper L, Gass D, Grol R, Mant D, Silagy C editor(s). *Oxford Textbook of Primary Medical Care*. Oxford (UK): Oxford University Press, 2004:134–8.

Slavin 2013
Slavin MJ, Brodaty H, Sachdev PS. Challenges of diagnosing dementia in the oldest old population. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 2013;**68**(9):1103–11.

StataCorp 2013 [Computer program]
StataCorp. Stata Statistical Software: Release 13. College Station, Tx: StataCorp LP, 2013.

van den Dungen 2012
van den Dungen P, van Marwijk HW, van der Horst HE, Moll van Charante EP, Macneil Vroomen J, van de Ven PM, et al. The accuracy of family physicians' dementia diagnoses at different stages of dementia: a systematic review. *International Journal of Geriatric Psychiatry* 2012;**27**(4):342–54.

Waldorff 2012
Waldorff FB, Siersma V, Vogel A, Waldemar G. Subjective memory complaints in general practice predicts future dementia: a 4-year follow-up study. *International Journal of Geriatric Psychiatry* 2012;**27**(11):1180–8.

Whiting 2011
Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529–36.

Whiting 2011a
Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *Journal of Clinical Epidemiology* 2011;**64**(6):602–7.

Winblad 2004
Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund L-O, et al. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine* 2004;**256**(3):240–6.

Woolley 2013
Woolley A, Kostopoulou O. Clinical intuition in family medicine: more than first impressions. *Annals of Family Medicine* 2013;**11**(1):60–6.

* *Indicates the major publication for the study*

ADDITIONAL TABLES

Table 1. Circumstances for contacting authors to obtain information on diagnostic accuracy

Aspect of study that is not relevant to our review	Action we will take
Participants	Exclude the study
Reference standard	Exclude the study
Index test	Exclude the study

Table 1. Circumstances for contacting authors to obtain information on diagnostic accuracy (*Continued*)

Target condition	Contact authors in the hope of obtaining information about diagnostic accuracy for target condition of interest only when we are confident from review of the full text that the participants, reference standard and index test are applicable to the review Where studies report the diagnostic accuracy of clinical judgement for the diagnosis of a composite target condition of cognitive impairment and dementia (e.g. cognitive impairment) we will attempt to obtain details of the diagnostic accuracy for each of our separate target conditions
Study design	Contact authors in the hope of obtaining information about diagnostic accuracy for target condition of interest only when we are confident from review of the full text that the participants, reference standard, index test and target condition are applicable to the review

APPENDICES

Appendix 1. MEDLINE search strategy

1. exp "sensitivity and specificity"/
2. "reproducibility of results"/
3. diagnos*.ti.
4. di.fs.
5. sensitivit*.ab.
6. specificit*.ab.
7. (ROC or "receiver operat*").ab.
8. Area under curve/
9. ("Area under curve" or AUC).ab.
10. sROC.ab.
11. accura*.ti,ab.
12. (likelihood adj3 (ratio* or function*)).ab.
13. ((true or false) adj3 (positive* or negative*)).ab.
14. ((positive* or negative* or false or true) adj3 rate*).ti,ab
15. or/1-14
16. exp Dementia/
17. Delirium, Dementia, Amnestic, Cognitive Disorders/
18. dement*.mp.
19. alzheimer*.mp.
20. (lewy* adj2 bod*).mp.
21. (chronic adj2 cerebrovascular).mp.
22. ("organic brain disease" or "organic brain syndrome").mp

23. ("normal pressure hydrocephalus" and "shunt*").mp.
24. "benign senescent forgetfulness".mp.
25. (cerebr* adj2 deteriorat*).mp.
26. (cerebral* adj2 insufficient*).mp.
27. (pick* adj2 disease).mp.
28. (creutzfeldt or jcd or cjd).mp.
29. huntington*.mp.
30. binswanger*.mp.
31. korsako*.mp.
32. "cognit* impair*".mp.
33. exp *Cognition Disorders/
34. MCI.ti,ab.
35. ACMI.ti,ab.
36. ARCD.ti,ab.
37. SMC.ti,ab.
38. CIND.ti,ab.
39. BSF.ti,ab.
40. AAMI.ti,ab.
41. MD.ti,ab.
42. LCD.ti,ab.
43. QD.ti,ab.
44. AACD.ti,ab.
45. MNCD.ti,ab.
46. MCD.ti,ab.
47. ("N-MCI" or "A-MCI" or "M-MCI").ti,ab.
48. ((cognit* or memory or cerebr* or mental*) adj3 (declin* or impair* or los* or deteriorat* or degenerat* or complain* or disturb* or disorder*)).ti,ab
49. "preclinical AD".mp.
50. "pre-clinical AD".mp.
51. ("preclinical alzheimer*" or "pre-clinical alzheimer*").mp
52. (aMCI or MCIa).ti,ab.
53. ("CDR 0.5" or "clinical dementia rating scale 0.5").ti,ab
54. ("GDS 3" or "stage 3 GDS").ti,ab.
55. ("global deterioration scale" and "stage 3").mp.
56. "mild neurocognit* disorder*".ti,ab.
57. (prodrom* adj2 dement*).ti,ab.
58. (episodic* adj2 memory).mp.
59. ("preclinical dementia" or "pre-clinical dementia").mp.
60. or/16-59
61. Family Practice/ or Ambulatory Care/
62. Physicians, Family/ or Physicians, Primary Care/
63. Primary Health Care/
64. "family practice".ti,ab.
65. "general practi*".ti,ab.
66. *General Practice/ or General Practitioners/
67. "family practices".ti,ab.
68. "family practitioner*".ti,ab.
69. "primary care".ti,ab.
70. Physician Assistants/

71. "physician assistant*".ti,ab.
72. Nurse Practitioners/ or Family Nurse Practitioners/
73. "nurse practitioner*".ti,ab.
74. or/61-73
75. 60 and 74
76. 15 and 75
77. "clinical judgement*".ti,ab.
78. "practitioner* judgement*".ti,ab.
79. ((clinician* or GP* or physician* or doctor*) adj3 (intuit* or recognis* or recogniz* or reason* or detect* or diagnos*)).ti,ab
80. "gut feeling*".ti,ab.
81. gestalt.ti,ab.
82. "GP judgement*".ti,ab.
83. ((clinician* or GP* or physician* or doctor*) adj3 accura*).ti,ab
84. *Practice Patterns, Physicians'/
85. or/77-84
86. 60 and 85
87. 77 or 86

Appendix 2. Anchoring statements for assessment of risk of bias using QUADAS -2

Selection	Index test	Reference standard	Flow
Was a consecutive or random sample of patients enrolled? [yes/no] Consecutive or random sampling from patients in primary care would be considered at low risk of bias	*Were the index test results interpreted without knowledge of the results of the reference standard? [yes/no]* Studies at low risk of bias are likely to use terms such as "blinded" or "masked". Studies that do not explicitly state that access to medical records was denied will be judged as unclear. See Index tests.	*Is the reference standard likely to correctly classify the target condition? [yes/no]* See Reference standards. We will only include studies that use a recognised research definition of dementia which we will judge at low risk of bias	*Was there an appropriate interval between index test(s) and reference standard? [yes/no]* A study with an average delay between assessments of six months or less would be judged at low risk of bias. A study with a average delay of more than a year would be judged at high risk of bias. For delayed follow up as a reference standard, follow up should occur at least three months after the index test assessment
Was a case-control design avoided? [yes/no] We will not include case-control studies.	*If a threshold was used, was it prespecified? [yes/no]* See Data extraction and management. There is no accepted cut point for the index test. This item is likely to be of limited value in this review	*Were the reference standard results interpreted without knowledge of the results of the index test? [yes/no]* Studies at low risk of bias are likely to use terms such as "blinded" or "masked". Studies that state that the reference standard assessment was	*Did all patients receive a reference standard? [yes/no]* Many studies in primary care that are not primarily designed as prospective research studies may be at high risk of bias in this domain. See Index tests.

(*Continued*)

		allowed knowledge of the index test will be judged as high risk. Many studies may be at unclear risk of bias in this domain because of the possibility of referral letters from GPs to specialists Cross-sectional studies may be at higher risk of bias in this domain unless masking is explicit	
Did the study avoid inappropriate exclusions? [yes/no] Example of high risk of bias would be exclusions based solely on age, educational attainment or place of residence. Example of low risk of bias would be terminally ill people			*Did all patients receive the same reference standard? [yes/no]* It is likely that at least some participants will not receive the reference standard in all studies
			Were all patients included in the analysis? *[yes/no]* A maximum proportion of drop outs to remain low risk of bias has been specified as 20%
Could the selection of patients have introduced bias? [High/low/unclear] If exclusions are not explicit in the article or after contacting authors we will judge this as unclear Studie at high risk of bias would often use a sampling method that is not consecutive or random and / or exclude people inappropriately	*Could the conduct or interpretation of the index test have introduced bias?* *[High/low/unclear]* see Index tests We propose that the core feature of clinical judgement is that it is unaided by any additional test, investigation or inquiry beyond that which is immediately available to the clinician. Provided that the index test meets the definition we use the risk of bias for this item may be low risk. However, if it is not explicit that no other brief cognitive tests were used then the item may be at unclear risk of bias	*Could the reference standard, its conduct, or its interpretation have introduced bias?* *[High/low/unclear]* Even allowing for an acceptable reference standard studies may often be at unclear risk of bias in this domain unless it is explicit that the reference standard was applied independently of the index test	*Could the patient flow have introduced bias? High/low/unclear]* Many studies that are not primarily designed as research studies are likely to be at high risk of bias in this domain
Are there concerns that the included patients do not match the review question?	*Are there concerns that the index test, its conduct, or interpretation differ from the review question?*	*Are there concerns that the target condition as defined by the reference standard does not match the*	

(*Continued*)

Studies with high applicability will commonly include frail elderly people with multi-morbidity. Studies with low applicability will exclude these people. Studies with a prevalence of dementia of more than 70% will often be of low applicability	see Index tests. So long as the clinical judgement about dementia has been made by a primary care physician / general practitioner we will judge this at high applicability	*review question?* So long as the reference standard is one of our listed definitions we will judge this at high applicability	

CONTRIBUTIONS OF AUTHORS

All authors contributed to the manuscript and approved the submitted version.

DECLARATIONS OF INTEREST

There are no interests to declare.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- NIHR, UK.

This protocol was supported by the National Institute for Health Research (NIHR), via Cochrane Infrastructure funding to the Cochrane Dementia and Cognitive Improvement group. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, National Health Service or the Department of Health.