Informant-based screening tools for dementia: an overview of systematic reviews

Martin Taylor-Rowan[1]; Sara Nafisi[1]; Rhiannon Owen[2]; Robyn Duffy[3;] Amit Patel[2]; Jennifer K Burton[1]; Terence J Quinn[1]

1. Institute of Cardiovascular and Medical Sciences; University of Glasgow, 126 University Place, Glasgow, G12 8TA, United Kingdom.
2. Swansea University Medical School, Swansea University, Singleton Park, Swansea, SA2 8PP, United Kingdom.
3. Older People's Psychology Service, NHS Greater Glasgow and Clyde, United Kingdom.

Martin.taylor-rowan@glasgow.ac.uk (corresponding author)

Terry.quinn@glasgow.ac.uk

Sara.nafisi@istinye.edu.tr

Jenni.burton@glasgow.ac.uk

A.patel.13@bham.ac.uk

r.k.owen@swansea.ac.uk

Robyn.duffy@ggc.scot.nhs.uk

Word Count: 3740

Figures: 1

Tables: 4

Supp materials: Figures 3, Tables 11

Abstract

Background: Informant-based questionnaires may have utility for cognitive impairment or dementia screening. Reviews describing accuracy of respective questionnaires are available, but their focus on individual questionnaires precludes comparisons across tools. We conducted an overview of systematic reviews to assess comparative accuracy of informant questionnaires and identify areas where evidence is lacking.

Methods: We searched 6 databases to identify systematic reviews describing diagnostic test accuracy of informant questionnaires for cognitive impairment or dementia. We pooled sensitivity and specificity data for each questionnaire and used network approaches to compare accuracy estimates across the differing tests. We used Grading of Recommendations, Assessment, Development and Evaluation (GRADE) to evaluate overall certainty of evidence. Finally, we created an evidence 'heat-map', describing availability of accuracy data for individual tests in differing populations and settings.

Results: We identified 25 reviews, consisting of 93 studies and 13 informant questionnaires. Pooled analysis (37 studies;11,052 participants) ranked the 8-item interview to Ascertain Dementia (AD8) highest for sensitivity (90%; 95%CrI=82%-95%; 'best-test' probability=36%); while the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) was most specific (81%; 95%CrI=66%-90%; 'best-test' probability=29%). GRADE-based evaluation of evidence suggested certainty was 'low' overall. Our heat-map indicated only AD8 and IQCODE have been extensively evaluated and most studies have been in the secondary care setting.

Conclusions:AD8 and IQCODE appear to be valid questionnaires for cognitive impairment or dementia assessment. Other available informant-based cognitive screening questionnaires lack evidence to justify their use at present. Evidence on accuracy of available tools in primary care settings and with specific populations is required.

Key words: Cognitive impairment; dementia; informant; screening; systematic review; overview; informant

Background

Various assessment tools are available for screening of cognitive impairment or dementia. The most commonly used tests directly assess cognition via questions or 'pencil and paper' tasks. (Harrison, Noel-Storr, Demeyere, Reyish, & Quinn, 2016) These direct assessments provide a 'snapshot' of cognitive function that does not capture change in cognition, yet cognitive deterioration is a fundamental component of dementia diagnosis. In addition, direct assessments are often compromised, or not possible, in various acute secondary care settings. (Elliott et al., 2019) There is a need, therefore, to identify measures that can provide an alternative to traditional 'direct' cognitive screening methods.

An attractive approach is to assess cognition using informant-based interview tools. Through this method, a patient's close relative or friend (i.e. informant) is used to indirectly identify temporal change in patients' cognition and related function.

There are several informant tools available that are used in practice, such as the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE), (Jorm and Jacomb, 1989) the 8-item interview to Ascertain Dementia (AD8), (Galvin et al., 2005) and the General Practitioner Assessment of Cognition (GPCOG). (Brodaty et al., 2002) Current guidelines recommend use of structured informant interviews for cognitive assessment, but do not recommend a particular tool in preference to others. (NICE, 2020)

A number of systematic reviews have attempted to establish the diagnostic accuracy of informant-based tools in order to inform best tool selection. (Quinn et al., 2014; Harrison et al., 2014; Harrison et al., 2015; Harrison et al., 2016) However, this rapidly growing literature

may be overwhelming for clinicians and decision-makers, and to date has only considered available tools in isolation, precluding an answer to the question: which tool is best?

Novel evidence synthesis techniques (Owen RK, Cooper NJ, Quinn TJ, Lees R, & Sutton, 2018) allow for comparative assessment and are well suited to analysis of the accuracy of the various informant tools. A synthesis of published systematic reviews, i.e. an overview of systematic reviews, combined with a comparative summary could help to concisely summarise the broader evidence-base, improving clinicians' and policy makers' ability to select or recommend tools for cognitive assessment.

Aims and objectives

We performed an overview of systematic reviews to draw together results from systematic reviews of the diagnostic properties of informant-based cognitive screening tools.

Our primary question was: what is the comparative accuracy of informant-based screening tools for identifying cognitive impairment or dementia?

Secondary objectives

Where possible, we used this overview of systematic reviews to inform a number of secondary objectives:

To determine variability in informant tool diagnostic test accuracy across various settings and cognitive syndromes.

To evaluate the quality of systematic reviews of diagnostic test accuracy research such that common methodological issues can be highlighted, and standards improved.

To produce an 'evidence map' that reveals gaps in the evidence-base where new primary research is needed.

Methods

Design

We used the PRISMA (preferred reporting for systematic review and meta-analysis) checklist for reporting in this overview of systematic reviews. (see supplemental materials e-1)

Design, conduct and interpretation of overviews of systematic reviews is evolving; we followed recent best practice guidance. (Higgins et al., 2019; McKenzie & Brennan, 2017)

All aspects of searching, data extraction and review assessment were performed by two reviewers independently, with recourse to a third arbitrator where disagreement could not be resolved.

A detailed description of our methodology can be seen in the previously published protocol. (Taylor-Rowan, Nafisi, Patel, Burton & Quinn, 2020) A summary of our methodology is provided in the sections below.

Inclusion and exclusion criteria

We included systematic reviews that investigated the diagnostic properties (test accuracy) of an informant-based cognitive screening tool. We included reviews conducted in any setting or patient population. We operationalised the settings in which informant tools are used as: secondary care, primary care, and community. We made no exclusions on the basis of methodological quality, use of best practice methods, or approach to data synthesis.

Reviews were excluded if they exclusively reported on the diagnostic test accuracy of telephone-based assessment, prognostic accuracy, or 'functional' informant tools that measure ability to perform activities of daily living, rather than cognition *per se*. We also excluded non-English language reviews.

Search methods for identification of reviews

We searched EMBASE (OVID); Health and Psychosocial Instruments (OVID); Medline (OVID); CINAHL (EBSCO); PSYCHinfo (EBSCO) and the PROSPERO registry of review protocols. All databases were searched from inception to December 2019. Search syntax can be seen in supplementary materials (e-2).

We additionally contacted authors working in the field of dementia test accuracy to identify other relevant systematic reviews, and studied reference lists of all included reviews in order to identify additional titles not found by our search. (Greenhalgh & Peacock, 2005)

Data collection and analysis

Title selection and data extraction

Titles were screened using Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia, available at www.covidence.org. Data was extracted on to a data collection proforma that was specifically designed by the author team (see supplementary materials; e-3)

Assessment of methodological and reporting quality of included reviews

Methodological quality of included reviews was evaluated using a modified version of the AMSTAR-2 (assessment of multiple systematic reviews) measurement tool (Shea et al., 2017) which considered the following key domains: clarity of review objective; description of study eligibility criteria; extent of searching undertaken; transparency of assessment process; assessment of publication bias; assessment of heterogeneity. Overall study quality conclusions were established based on guidance from Shea et al. (2017). However, as this guidance is based on reviews of healthcare interventions, we modified the critical domains to include only: adequacy of the literature search (item 4); risk of bias from individual studies included in the review (item 9); appropriateness of meta-analytical methods (item 11); and consideration of risk of bias when interpreting the results of the review (item 13). (see supplementary materials; e-4)

AMSTAR-2 assessment was complimented with an evaluation of reporting standards of included reviews, utilising the PRISMA-DTA (Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies) checklist. (McInnes et al., 2018)

Data synthesis

We extracted data for analyses directly from original papers identified within respective reviews. We calculated summary estimates for each informant questionnaire using the bivariate approach (Reitsma, Glas, Scholten, Bossuyt & Zwinderman, 2005). Where suitable data (defined below) were available, we then conducted comparative analyses, creating a network where each questionnaire at a particular threshold score is a node and inferences around relative test performance can be made through indirect comparison and ranking. We used a bivariate network meta-analysis model accounting for the correlations between multiple test accuracy measures from the same study. (Owen et al., 2018; O'Sullivan, 2019) All models were estimated in a Bayesian framework using Markov Chain Monte Carlo (MCMC) simulation and implemented in the WinBUGS 1.4.3 software. (Lunn, Thomas, Best, & Spiegelhalter., 2000) Non-informative prior distributions were specified for test and threshold-specific accuracy parameters. Informant-based screening tools with the highest sensitivity and specificity were ranked in first place at each MCMC iteration. The estimated rankings overall were calculated as a summary of the individual ranks at each iteration. The probability that each screening tool was the best overall was calculated as the proportion of MCMC iterations that each informant tool ranked in first place. Further details on the analyses used are available in the original paper describing the method. (Owen et al., 2018)

We only included studies that evaluated informant tool test accuracy against a diagnostic standard consistent with recognised criteria for diagnosis of dementia or MCI (e.g. ICD-10, DSM III-V). We attempted meta-analysis where informant tools were assessed in at least two studies. Case-control studies were excluded due to the potential to over inflate test accuracy. For our primary analysis, we restricted analysis to the cut-points that were most regularly used and of most clinical relevance (3.3. and 3.6 for IQCODE; 2 & 3 for AD8). As

our primary question was to evaluate the accuracy of tools as measures of cognitive impairment or dementia (all inclusive), we did not discriminate between forms of cognitive impairment evaluated in included studies. However, where single studies provided sensitivity and specificity data for multiple forms of cognitive screening (e.g. sensitivity/specificity values for screening of dementia vs no dementia and sensitivity/specificity values for screening 'any cognitive impairment' vs normal cognition), we selected one reported sensitivity and specificity figure based on the following hierarchy: 'any cognitive impairment vs normal cognition'> 'dementia vs no dementia'> 'Mild Cognitive Impairment' (MCI)vs normal cognition'.

We employed GRADE (Grading of Recommendations Assessment, Development, and Evaluation) (Guyatt et al., 2008) to evaluate overall strength of sensitivity and specificity evidence for each tool in our meta-analysis, following recommended guidelines on application of GRADE to diagnostic test accuracy evidence. (Singh, Chang, Matchar & Bass., 2012)

Subgroup analysis

In addition to our primary analysis, we conducted a subgroup analyses designed to provide specific data on performance of tools when used to screen for cognitive syndromes of differing severity and when used in particular settings. Specifically, we evaluated performance of respective informant tools when used to differentiate between people with and without dementia (dementia vs no dementia) and between people with MCI and normal cognition (MCI vs normal cognition). For each analysis, we sub-grouped by setting (primary care, secondary care and community care), where possible.

Sensitivity analysis

We conducted a sensitivity analysis restricting to studies that had no high risk of bias categories and at least 50% low risk of bias categories (based on individual study level data within the included review).

Method for generation of evidence map

In addition to our search for relevant reviews, we identified individual (i.e. non-review) informant-based diagnostic test accuracy studies to generate an 'evidence heat-map'.

Search strategy for evidence map

We accessed referenced studies in included reviews and supplemented this with a search of study reference lists and, where provided, review exclusion lists for further available studies.

Inclusion/exclusion criteria for evidence map

To be included in the evidence heat-map, individual studies could be either cohort or case-control, but were required to be published in a peer-reviewed scientific journal and report on the diagnostic test accuracy (i.e. sensitivity and specificity) of an informant tool. We included non-English language papers in our evidence heat-map, but studies were excluded if they reported participant numbers <20; were abstracts; were repeat data sets; assessed prognostic diagnostic test accuracy; described a 'functional' informant measure only (e.g.

Independent activities of daily living scale); or if the informant tool was completed by patients rather than informants.

Extent of available evidence was depicted via a shading scheme ranging from dark (0-10 studies; limited evidence), to light (>40 studies; substantial evidence).

Results

Our search identified 4865 titles. After screening, we found 25 reviews (including 93 studies) that met our inclusion criteria. (see Table 1) Details of the screening process and reasons for each exclusion can be seen in supplementary materials (e-5).

[insert Table 1]

Summary of reviews' findings

Thirteen informant-based assessment tools were discussed in included reviews. The diagnostic test accuracy properties of 11 of these tools were described. Each reviewed tool is presented below.

IQCODE

The most comprehensively assessed informant tool was the IQCODE, which was included in 18 reviews and 52 original studies. Five distinct versions of the IQCODE were described based on the number of component question items (IQCODE-32, IQCODE-26, IQCODE-16, IQCODE-17, IQCODE-7); the most commonly used versions were the 26-item and the 16-item adaptation.

Pooled estimates of IQCODE accuracy for dementia diagnosis ranged from sensitivity 80-91% and specificity 66-85%. Review evaluations of IQCODE diagnostic test accuracy studies suggested study quality was generally poor. In Cochrane reviews, (Quinn et al., 2014; Harrison et al., 2014; Harrison et al., 2015) just 2/25 IQCODE studies were judged to have no high risk of bias categories. Typical issues were around lack of blinding and unnecessary patient exclusions—particularly removal of those who may benefit most from an informant-based assessment (e.g. patients with comorbidities that make traditional cognitive assessments challenging).

AD8

The AD8 was assessed in 5 reviews (20 studies). Pooled sensitivity rates for dementia diagnosis ranged from 88-97% and pooled specificity rates ranged from 64-81%. Cochrane review evaluations (Hendry et al., 2019) determined that 4/10 AD8 studies had no high risk of bias categories. Areas of study limitation were around inadequate reporting, inappropriate exclusions of participants, and high participant drop-out rates due to inability to complete tests.

GPCOG

The GPCOG was evaluated in 6 reviews, describing 5 distinct studies.

All but two reviews evaluated the diagnostic test accuracy of the GPCOG based on the evidence of just 1 'fair quality' (Lin, O'Connor, Rossom, Perdu & Eckstrom., 2013) study. A more recent review (Tsoi, Chan, Hirai, Wong & Kwok., 2015) evaluated 5 GPCOG studies and reported a pooled sensitivity of 92% and specificity of 87%. However, risk of bias was substantial (25% of studies rated high risk of bias in 3 out of 4 domains). Unlike most other

299 informant tools, the GPCOG has a combined patient and informant assessment. When the
300 informant component of the GPCOG was used in isolation, it appeared to have poor
301 specificity (49-66%). (Kansagara & Freeman., 2010)

302 **Other informant-based assessment tools**

303 Ten additional informant tools were described in at least one included review. A summary
304 of the diagnostic test accuracy evidence for each can be seen in Table 2.

305 [insert Table 2]

306 **Network meta-analysis**

307 From each review, we identified a total of 37 suitable studies (11,052 participants) to
308 evaluate comparative performance of respective tools. One study (Jorm et al., 1996)
309 provided direct (within study) comparative data on the IQCODE-26 and IQCODE-16; 2
310 studies (Jackson, MacLullich, Gladman, Lord & Sheehan, 2016; Razavi et al., 2014) provided
311 direct comparative data on IQCODE-16 and AD8. All other studies provided test accuracy
312 properties of single informant tools in isolation, meaning indirect (between study)
313 comparisons were predominant in our network meta-analyses.

314 **Primary analysis**

315 Our primary network meta-analysis examined performance of informant tools as measures
316 of cognitive impairment or dementia (all inclusive). Only 3 informant tools had sufficient
317 data for comparative analysis (IQCODE-26; IQCODE-16 & AD8).

318 Results suggest AD8 at cut-point 2 may have the highest sensitivity (90%; 95% credible
319 intervals [CrI]=82%-95%; 'best test' probability=36%) for detecting cognitive impairment or
320 dementia, although there was little difference between AD8 at cut point 2, AD8 at cut point

3 and IQCODE-16 at cut point 3.6 with probability best of 36%, 23%, and 22% respectively. IQCODE-26 at cut-point 3.6 may have the highest specificity (81%; 95%CrI=66%-90%; 'best test' probability= 29%), though again there was little difference between IQCODE-26 at cut-point 3.6, IQCODE-16 at cut point 3.6, and IQCODE-16 at cut point 3.3 with probability best of 29%. 26% and 17%, respectively. We noted that two studies (Jackson, MacLullich, Gladman, Lord & Sheehan, 2016; de Jonghe, 1997) were conducted in distinct populations (delirious and depressed, respectively) that could alter diagnostic test accuracy properties. We therefore conducted an additional sensitivity analysis, removing these 2 studies. Results were unchanged. (see supplementary materials; e-6)

Comparative performance for each tool at respective cut-points can be seen in Table 3.

[insert Table 3]

Subgroup analysis

We evaluated the performance of tools when screening for a specific cognitive syndrome in a particular setting. Sufficient data for pooling in this subgroup analysis was only available for respective tools at certain cut-points. (see Table 4)

Comparative data on tool performance for 'dementia vs no dementia' screening suggests that the AD8 at cut-point 2 may have the highest sensitivity for dementia in both secondary care (96%; 95%CrI=72-99%; 'best test' probability= 76%) and community settings (86%; 95%CrI=64-95%; 'best test' probability=48%). IQCODE-16 at cut point 3.3 had the greatest specificity for dementia assessment in secondary care (71%; 95%CrI=35-93%; 'best test' probability=73%) while IQCODE-26 at cut-point 3.6 had the highest specificity (93%; 95%CrI=81-98%%; 'best test' probability=90%) in the community.

Comparisons of general tool performance across settings suggest sensitivity of each tool is consistently higher when used in the secondary care setting than when used in the community (secondary care sensitivity range: 82-96%; community care sensitivity range: 68-86%), whereas specificity is comparatively reduced (secondary care specificity range: 39-71%; community care specificity range:71-93%).

[insert Table 4]

There were insufficient studies to compare tool performance when used in primary care or for assessing MCI vs normal cognition.

Risk of Bias sensitivity analysis

We evaluated reported rates when restricted to studies deemed to be at lower risk of bias. Seven studies were available in total; however, there was too much heterogeneity to pool data, hence individual study findings were assessed. (Supplementary materials, e-6) The general trend of informant tool performance was consistent with our pooled analyses.

Strength of overall evidence

Our GRADE rating of the strength of the IQCODE and AD8 diagnostic test accuracy evidence was 'low' for sensitivity and specificity of both tools, primarily due to the risk of bias present in included studies and the imprecision apparent in our pooled rates. (see supplementary materials, e-7)

Overview of systematic reviews—evaluation of review methodological and reporting quality

Our AMSTAR-2 evaluations highlighted a number of methodological issues in included reviews. Overall review quality was mixed: 8/25 (32%) reviews were 'critically low' quality; 6/25 (24%) reviews were rated moderate and 3/25 (12%) were high methodological quality. All reviews rated moderate or above were conducted from 2010 onwards (see supplemental materials for AMSTAR-2 evaluation, e-8). All reviews performed a comprehensive search and study inclusion criteria was generally adequately explained. However, a number of reviews did not perform the systematic search and/or conduct data-extraction in duplicate via 2 independent investigators (9/25; 36%); errors in data extraction were frequent, and very few reviews pre-registered a protocol (5/25; 20%).

Meta-analyses were performed in 11/25 (44%) reviews and appropriate statistical methods were used in each—though it was common for reviews to include case-control studies in pooled analyses, potentially exaggerating diagnostic test accuracy. (Higgins et al, 2019)

Risk of bias was not adequately investigated in 9/25 (36%) reviews. Where risk of bias assessment was conducted, conclusions regarding individual studies were often contrasting. For instance, Chen et al. (2017) rated all seven included AD8 studies to be 'high quality', identifying no high risk of bias domains in any study; Hendry et al. (2019) rated 4/7 of the same studies to have at least 1 high risk of bias domain. No reviews conducted a sensitivity analysis gauging the impact of high risk of bias studies upon reported pooled results, and only 1 review (Chen et al., 2017) investigated possible publication bias.

Evaluation of reporting standards via PRISMA-DTA revealed main issues around explicit statements of objectives (12/25 [48%] studies), describing information sources in adequate

detail (12/25 [48%] studies) and reporting sufficient details of test accuracy from individual included studies (11/25 [44%] studies).

Evidence Map findings

A total of 93 distinct informant tool studies were identified and diagnostic test accuracy properties were described across a range of settings and populations. (Figure 1) Our findings suggests that IQCODE and AD8 have a greater evidence-base than other available tools, but there are a lack of diagnostic test accuracy evaluations in primary care and specialised populations (e.g. stroke). References of included papers, along with risk of bias judgements for each included study can be seen in supplementary materials (e-9).

[insert Figure 1: evidence map]

Discussion

Comparative evidence for available tools

At least 13 informant tools for cognitive assessment are available, though there is a lack of evidence to justify use of all but two of these tools: the IQCODE and the AD8. The reviewed literature suggests that both tools have reasonable diagnostic test accuracy for assessment of cognitive impairment or dementia, comparable with other popular cognitive screening tools such as the Mini Mental State Examination and Montreal Cognitive Assessment. (Tsoi,

et al., 2015) Our network meta-analysis indicates the AD8 may be the more sensitive of the two tools, and the IQCODE the more specific; however, the credible intervals (CrI) were overlapping and estimates of 'best test' probability were close for both sensitivity and specificity, implying little performance difference between respective tools. The overall strength of the available evidence was also low according to our GRADE evaluation, tempering conclusions.

Our findings highlight that the general performance of each tool is variable and typically lower than originally suggested by the developers. (Jorm & Jacomb, 1989; Galvin et al., 2005) Moreover, while both tools appear capable of screening for dementia, test performance may vary by setting. When used in specialised secondary care settings, where specificity may be the preferred property, at traditional clinical thresholds neither tool appears well-suited to differentiating patients with dementia from those with mild or age-related cognitive changes. Though the IQCODE-16 demonstrated a reasonable specificity of 73% in secondary care at cut point 3.3, this value was inconsistent with the suggested performance (57%) of the longer IQCODE-26 at a cut point (3.6) that prioritises specificity; thus, this may be an example of study bias exaggerating tool performance. Specificity may be comparatively higher in community settings. However, in this setting, sensitivity may be the preferred property.

We therefore suggest that neither informant tool is well suited for use as a solitary cognitive screening tool. However, these tools can still be useful as solitary assessments in instances where patients are unable or unwilling to complete a more direct test; thus, where clinicians seek to employ an informant tool, selection of the IQCODE or AD8 should be guided by

desire for sensitivity or specificity. The AD8 at cut point 2 will likely provide the greatest sensitivity, while the IQCODE-26 at cut point 3.6 will provide the greatest specificity.

It is important to emphasise that our analyses were designed to assess test accuracy only. Other properties are also important for consideration when selecting an appropriate tool for cognitive screening. Feasibility, inter-rater reliability, responsiveness to change, and suitability for use in specialist populations are all important test characteristics that may influence the selection of one test over another in clinical practice. While it is beyond the scope of this review to discuss each respective tool in these terms, we encourage further work on this topic to supplement the test accuracy finding we present here.

The state of diagnostic test accuracy literature

Previous overviews of systematic reviews have highlighted significant issues with regards to review methodological quality. (Arevalo-Rodriguez et al., 2014) We similarly found prevalent methodological issues, but also some promising signs.

In contrast to previous diagnostic test accuracy overviews of systematic reviews, the majority of our included reviews conducted formal risk of bias assessments and the higher quality reviews were all conducted within the previous decade, suggesting increasing standards.

However, that risk of bias assessments were inconsistent across reviews indicates a poor understanding of the ways in which a diagnostic test accuracy study design can introduce bias. Existing risk of bias assessment tools typically require investigators to tailor presented questions to the topic of interest. The robustness of this modification process is heavily

impacted by the amount of experience investigators have in the topic area; thus, subjectivity influences the process of assessing risk of bias even when formal rating tools are operationalised. Furthermore, study bias is generally under-considered when results are discussed: conclusions and recommendations are frequently made in reviews without full exploration of the potential impact biased studies may have had on pooled results. Clinicians should be mindful of these limitations when consuming the evidence provided in a review.

Gaps in the evidence-base

Our evidence map highlights the main areas in which informant tool test accuracy studies are a priority. Primary care has comparatively little evidence to other healthcare settings despite being arguably the most important location for cognitive screening or triage. (Quinn et al., 2014) Similarly, informant tool diagnostic test accuracy evaluations are lacking in specialised populations that typically struggle with more traditional cognitive tests (e.g. stroke populations). We would therefore encourage further work to determine the accuracy of available informant tools in these populations.

Future directions

While our data suggest that informant tools may not generally be suitable as solitary screening tools, they may have utility when combined with direct screening tests. Most available evidence suggests that direct and informant tools perform better when used

together. (e.g. Tew, Ng, Cheong, & Yap, 2015; Srikanth et al., 2006; Narasimhalu, Lee, Auchus, & Chen, 2008) Thus, informant tools may make ideal supplements to the standard cognitive assessment, yet no reviews exist on this topic.

This type of evaluation is very much needed if we are to confirm the value of a dual (i.e. direct and informant) approach to assessment. It is important to note that available tests (both direct and informant) typically cover varying cognitive domains; (Cullen et al., 2007) hence, the best combinations of tests may change dependent upon the types of cognitive problems that are present in a given population.

Strengths and limitations

We have conducted a comprehensive overview of systematic reviews that brings together the findings of 25 distinct reviews, depicts an extensive evidence map, and employs new statistical techniques that allow formal statistical comparisons, ranking, and 'best test' probability estimates between informant tools—addressing a major limitation of this literature.

However, our overview of systematic reviews has some limitations. Firstly, the credible intervals in our network meta-analysis are wide for our specificity estimates and most included studies are at risk of bias; hence, resultant rankings should not be viewed as definitive and uncertainty in these estimates should be considered.

Secondly, our comparisons between tools are overwhelmingly based on indirect comparisons, reliant upon statistical control for random variations in populations—although our findings are strengthened by a consistency with those studies that directly compared

the IQCODE and AD8 within the same participant pool. (Jackson, et al., 2016; Razavi et al., 2014).

Thirdly, due to limited study numbers, we were unable to conduct some of our pre-specified analyses, such as evaluations of tool performance in primary care settings.

Lastly, our evidence map is restricted to studies referenced in published systematic reviews; thus, there are some recently published studies and informant tools which have not been reviewed, such as the recently developed Quick Dementia Rating System (Galvin, 2015), that do not feature.

Conclusion

Our findings suggest that only the IQCODE and AD8 have had their diagnostic test accuracy properties widely evaluated. Based on available data, the AD8 at cut point 2 may be the most sensitive available tool for detecting cognitive impairment or dementia, while the IQCODE-26 at cut point 3.6 is the most specific. However, there is little evidence to suggest an important difference in tool performance overall, and neither tool performs well enough to be used alone for dementia assessment. Further evaluations of test accuracy in primary care and specialised populations are a priority.

Required Statements

Ethical approval: Ethics approval and consent to participate not required.

Consent for publication: All authors have seen the materials and consent.

Conflict of interests: Dr Owen is a member of the NICE Technology Appraisals Committee and the NICE Decision Support Unit (DSU). Dr Owen has served as a paid consultant to the pharmaceutical industry, not in relation to this research.

Funding: This work is funded by the National Institute of Health Research. The funders played no part in the conduct of this review.

Authors' contributions: TQ conceived the idea. MT and TQ designed the study and drafted the manuscript. SN and RD were the 2nd and 3rd reviewers on the paper. JB dealt with disagreements between reviewers. RO performed statistical analysis for the review. AP contributed to data interpretation and writing. MT is the Guarantor and all authors have read and commented on the final draft.

Acknowledgements: Thanks to Cochrane Test Accuracy Methods Group; and contributors to CRSU workshops.

References

Arevalo-Rodriguez I., Segura O., Sola I., Bonfill X., Sanchez E., & Alonso-Coello P. (2014) Diagnostic tools for alzheimer's disease dementia and other dementias: an overview of diagnostic test accuracy (DTA) systematic reviews. *BMC Neurology*, *14.* doi: 10.1186/s12883-014-0183-2.

Breton A., Casey D., & Arnaoutoglou N.A. (2019) Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of diagnostic accuracy studies. *International Journal of Geriatric Psychiatry, 34*(2), 233-242. doi: 10.1002/gps.5016.

Brodaty H., Pond D., Kemp N.M., Luscombe G., Harding L., Berman K., & Huppert F.A. (2002) The GPCOG: a new screening test for dementia designed for general practice. *Journal of the American Geriatrics Society*, *50*(3), 530-534. doi: 10.1046/j.1532-5415.2002.50122.x.

Burton L., & Tyson S.F. (2015) Screening for cognitive impairment after stroke: a systematic review of psychometric properties and clinical utility. *Journal of Rehabilitation Medicine*, *47*(3), 193–203. doi: 10.2340/16501977-1930.

Carpenter C.R., Banerjee J., Keyes D., Eagles D., Schnitker L., Barbic D., … LaMantia MA. (2019) Accuracy of Dementia Screening Instruments in Emergency Medicine: A Diagnostic Meta-analysis. *Academic Emergency Medicine, 26*(2), 226-245. doi: 10.1111/acem.13573.

Chen H.H., Sun F.J., Yeh T.L., Liu H.E., Huang H.L., Kuo B.I.T., & Huang H.Y. (2017) The diagnostic accuracy of the Ascertain Dementia 8 questionnaire for detecting cognitive impairment in primary care in the community, clinics and hospitals: a systematic review and meta-analysis. *Family Practice*, 35(3), 239-246. doi:10.1093/fampra/cmx098.

Cherbuin N., Antsey K.J., & Lipnicki D.M. (2008) Screening for dementia: a review of self- and informant-assessment instruments. *International Psychogeriatrics*, *20*(3), 431–458. doi:10.1017/S104161020800673X.

Cullen B., O'Neill B., Evans J.J., Coen R.F., & Lawlor B.A. (2007) A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery and Psychiatry*, *78*(8), 790–799. doi: 10.1136/jnnp.2006.095414.

American Psychiatric Association. (2000) Diagnostic and statistical manual of mental disorders: DSM-IV-TR. Washington, DC.

de Jonghe, J.F.M. (1997) Differentiating between demented and psychiatric patients with the Dutch version of the IQCODE. *International Journal of Geriatric Psychiatry*, *12*(4), 462–465. doi: 10.1002/(sici)1099-1166(199704)12:4<462::aid-gps510>3.0.co;2-q.

Elliott, E., Drozdowska, B. A., Taylor-Rowan M., Shaw R. C., Cuthbertson G., & Quinn T. J. (2019). Who Is Classified as Untestable on Brief Cognitive Screens in an Acute Stroke Setting? *Diagnostics, 9*(3), 95. doi.org/10.3390/diagnostics9030095

Galvin J.E., Roe C.M., Powlishta K.K., Coats M.A., Much S.J., Grant E., … Morris J.C. (2005) The AD8: a brief informant interview to detect dementia. *Neurology, 65*(4), 559-564. doi: 10.1212/01.wnl.0000172958.95282.2a.

Galvin J.E. (2015). The Quick Dementia Rating System (QDRS): A rapid dementia staging tool. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(2), 249-259. doi: 10.1016/j.dadm.2015.03.003.

Greenhalgh T., & Peacock R. (2005) Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*, *331*(7524), 1064-1065. doi: 10.1136/bmj.38636.593461.68.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., Schunemann H.J. (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* (Clinical research ed), *336*(7650), 924-926. doi: 10.1136/bmj.39489.470347.AD

Harrison J.K., Noel-Storr A.H., Demeyere N., Reyish E.L., & Quinn T.J. (2016) Outcomes measures in a decade of dementia and mild cognitive impairment trials. *Alzheimer's Research & Therapy*, *8*(48). Doi: 10.1186/s13195-016-0216-8

Harrison J.K., Stott D.J., McShane R., Noel-Storr A.H., Swann-Price R.S., & Quinn T.J. (2016) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings. *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD011333

Harrison J.K., Fearon P., Noel-Storr A.H., McShane R., Stott D.J., & Quinn T.J. (2014) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting. *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD010771

Harrison J.K., Fearon P., Noel-Storr A.H., McShane R., Stott D.J., & Quinn T.J. (2015) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD010772

Harvan J.R., & Cotter V.T. An evaluation of dementia screening in the primary care setting. (2006) *Journal of the American Academy of Nurse Practitioners, 18*(8):351-60. doi: 10.1111/j.1745-7599.2006.00137.x

Hendry K., Green C., McShane R., Noel-Storr A.H., Stott D.J., Anwer S., … Quinn T.J. (2019) AD-8 for detection of dementia across a variety of healthcare settings*. Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD011121.pub2.

Hendry K., Hill E., Quinn T.J., Evans J., & Stott D.J. (2015) Single screening questions for cognitive impairment in older people: a systematic review. *Age and Ageing, 44*(2), 322–326. doi: 10.1093/ageing/afu167.

Higgins J.P.T., Thomas J., Chandler J., Cumpston M. Li T., Page M.J., & Welch V.A. (2019) Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). *Cochrane Database of Systematic Reviews*. doi:10.1002/9781119536604

Jackson T.A., MacLullich A.M., Gladman J.R., Lord J.M., & Sheehan B. (2016) Diagnostic test accuracy of informant-based tools to diagnose dementia in older hospital patients with delirium: a prospective cohort study. *Age and Ageing*, *45*(4), 505-511. doi: 10.1093/ageing/afw065

Jackson T.A., & Naqvi S.H., Sheehan B. (2013) Screening for dementia in general hospital inpatients: a systematic review and meta-analysis of available instruments. *Age and Ageing*, *42*(6), 689–695. doi: 10.1093/ageing/aft145.

Jorm A.F., & Jacomb P.A. (1989) The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine*, *19(4)*, 1015-1022. doi: 10.1017/s0033291700005742

Jorm A.F., Broe G.A., Creasey H., Sulway M.R., Dent O., Fairley M.J., …Tennant C. (1996) Further data on the validity of the informant questionnaire on cognitive decline in the elderly (IQCODE). *International Journal of Geriatric Psychiatry*, *11*(2), 131-139.

Jorm A.F. (1997) Methods of screening for dementia: a meta-analysis of studies comparing an informant questionnaire with a brief cognitive test. *Alzheimer Disease and Associated Disorders*, *11*(3), 158-162.

Jorm A.F. (2004) The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): a review. *International Psychogeriatrics*, *16*(3), 1–19. doi: 10.1017/S1041610204000390.

Kansagara D., & Freeman, M.A. (2010) Systematic Evidence Review of the Signs and Symptoms of Dementia and Brief Cognitive Tests Available in VA [Internet]. *Washington (DC): Department of Veterans Affairs (US).* Retrieved from https://www.hsrd.research.va.gov/publications/esp/dementia-REPORT.pdf

Kosgallana A., Cordato D., Chan D.K.Y., & Yong J. (2019) Use of Cognitive Screening Tools to Detect Cognitive Impairment After an Ischaemic Stroke: a Systematic Review. *SN Comprehensive Clinical Medicine* 1, 255–262. https://doi.org/10.1007/s42399-018-0035-2

Lin J.S., O'Connor E., Rossom R.C., Perdu L.A., & Eckstrom E. (2013) Screening for Cognitive Impairment in Older Adults: A Systematic Review for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, *159*(9), 601-612.

Lischka A.R., Mendlesohn M., Overend T., & Forbes D. (2012) A Systematic Review of Screening Tools for Predicting the Development of Dementia. *Canadian Journal on Aging*, *31*(3), 295– 311. doi:10.1017/S0714980812000220.

Lunn D.J., Thomas A., Best N., & Spiegelhalter D. (2000) WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337. doi.org/10.1023/A:1008929526011

McGovern A., Pendlebury S.T., Mishra N.K., Fan Y., & Quinn T.J. (2016) Test Accuracy of Informant-Based Cognitive Screening Tests for Diagnosis of Dementia and Multidomain Cognitive Impairment in Stroke. *Stroke*. 47(2), 329-335. doi: 10.1161/ STROKEAHA.115.011218.

McInnes M.D.F., Moher D., Thombs B.D., McGrath T.A., Bossuyt P.M., & Grp P.D. (2018) Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies The PRISMA-DTA Statement. *Journal of the American Medical Association*. 319(4), 388-396. doi: 10.1001/jama.2017.19163

McKenzie J.E., Brennan S.E. (2017) Overviews of systematic reviews: great promise, greater challenge. *Systematic Reviews*, *6*, 185. doi.org/10.1186/s13643-017-0582-8

Narasimhalu K., Lee J., Auchus A.P., & Chen C.P. (2008) Improving detection of dementia in Asian patients with low education: combining the Mini-Mental State Examination and the Informant Questionnaire on Cognitive Decline in the Elderly. *Dementia and Geriatric Cognitive Disorders*. *25(1)*, 17–22 . doi: 10.1159/000111128.

NICE. (2020). Retrieved from http://pathways.nice.org.uk/pathways/dementia

O'Sullivan J.W. (2019) Network meta-analysis for diagnostic tests. *BMJ Evidence-Based Medicine*, *24*, 192-193. Retrieved from https://ebm.bmj.com/content/24/5/192.

Owen R.K., Cooper N.J., Quinn T.J., Lees R., & Sutton A.J. (2018) Network meta-analysis of diagnostic test accuracy studies identifies and ranks the optimal diagnostic tests and thresholds for healthcare policy and decision making. *Journal of Clinical Epidemiology*, *99*, 64-74. doi: 10.1016/j.jclinepi.2018.03.005

Quinn T.J., Fearon P., Noel-Storr A.H., Young C., McShane R., & Stott D.J. (2014) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD010079

Razack M.A.A, Ahmad N.A., Chan Y.Y., Kasim N.M., Yusof M., Abdul Ghani M.K.A. … Jamaluddin R. (2019) Validity of Screening Tools for Dementia and Mild Cognitive Impairment Among the Elderly in Primary Health Care: A Systematic Review. *Public Health*, *169*, 84-92. doi: 10.1016/j.puhe.2019.01.001.

Razavi M., Tolea M.I., Margrett J., Martin P., Oakland A., Tscholl D.W., … Galvin J.E. (2014) Comparison of 2 Informant Questionnaire Screening Tools for Dementia and Mild Cognitive Impairment AD8 and IQCODE. *Alzheimer Disease & Associated Disorders*, *28*, 156-161. doi: 10.1097/WAD.0000000000000008.

Reitsma J.B., Glas A.S., Rutjes A.W., Scholten R.J., Bossuyt P.M., Zwinderman A.H. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, *58(10)*, 982-990. doi: 10.1016/j.jclinepi.2005.02.022.

Rosli R., Tan M.P., Gray W.K., Subramanian P., & Chin A.V. (2016) Cognitive assessment tools in Asia: a systematic review. *International Psychogeriatrics*, *28*(2), 189–210. doi:10.1017/S1041610215001635.

Shea B.J., Reeves B.C., Wells G., Thuku M., Hamel C., Moran J., … Henry D.A. (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358. doi: 10.1136/bmj.j4008.

Singh S., Chang S.M., Matchar S.M., & Bass E.B. (2012) Chapter 7: Grading a Body of Evidence on Diagnostic Tests. *Journal of General Internal Medicine*, *27*, S47-55. doi: 10.1007/s11606-012-2021-9.

Srikanth V., Thrift A.G., Fryer J.L., Saling M.M., Dewey H.M., Sturm J.W., & Donnan G.A. (2006) The validity of brief screening cognitive instruments in the diagnosis of cognitive impairment and dementia after first-ever stroke. *International Psychogeriatrics*, *18*, 295–305. doi: 10.1017/S1041610205002711.

Taylor-Rowan M., Nafisi S., Patel A., Burton J.K., & Quinn T.J. (2020) Informant based screening tools for diagnosis of dementia, an overview of systematic reviews of test accuracy studies- Protocol. *Systematic Reviews, 9*, 271. doi: 10.1186/s13643-020-01530-3

Tew C.W., Ng T.P., Cheong C.Y., & Yap P. (2015) A brief dementia test with subjective and objective measures. *Dementia and Geriatric Cognitive Disorders Extra, 5*(3), 341–349. doi: 10.1159/000438455

Tsoi K.K.F, Chan J.Y.C., Hirai H.W., Wong SYS, & Kwok TCY. (2015) Cognitive Tests to Detect Dementia A Systematic Review and Meta-analysis*. JAMA Internal Medicine*, *175*(9), 1450-1458. doi:10.1001/jamainternmed.2015.2152.

Tsoi K.K.F., Chan J.Y.C., Hirai H.W., Wong A., Mok V.C.T., Lam LCW, … Wong S.Y.S. (2017) Recall Tests Are Effective to Detect Mild Cognitive Impairment: A Systematic Review and Meta-analysis of 108 Diagnostic Studies. *JAMDA 18(9)*, 807. doi: 10.1016/j.jamda.2017.05.016

Woodford H.J., & George J. (2007) Cognitive Assessment in the Elderly: A Review of Clinical Method. *Quarterly Journal of Medicine*, *100*(8), 469–484. doi:10.1093/qjmed/hcm051.

World Health Organization. (2004) ICD-10: international statistical classification of diseases and related health problems: tenth revision, 2nd ed. Retrieved from https://www.cdc.gov/nchs/icd/icd10.htm#:~:text=The%20International%20Classification%20of%20Diseases,death%20on%20the%20death%20certificate.