

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/149569/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Beishon, Lucy C., Elliott, Emma, Hietamies, Tuuli M., McArdle, Riona, O'Mahony, Aoife, Elliott, Amy R. and Quinn, Terry J. 2022. Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia. Cochrane Library 2022 (4) , CD013724. 10.1002/14651858.CD013724.pub2 file

Publishers page: <http://dx.doi.org/10.1002/14651858.CD013724.pub2>
<<http://dx.doi.org/10.1002/14651858.CD013724.pub2>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.
See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





Cochrane
Library

Cochrane Database of Systematic Reviews

Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia (Review)

Beishon LC, Elliott E, Hietamies TM, Mc Ardle R, O'Mahony A, Elliott AR, Quinn TJ

Beishon LC, Elliott E, Hietamies TM, Mc Ardle R, O'Mahony A, Elliott AR, Quinn TJ.
Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia.
Cochrane Database of Systematic Reviews 2022, Issue 4. Art. No.: CD013724.
DOI: [10.1002/14651858.CD013724.pub2](https://doi.org/10.1002/14651858.CD013724.pub2).

www.cochranelibrary.com

Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

WILEY

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	7
OBJECTIVES	8
METHODS	8
RESULTS	11
Figure 1.	12
Figure 2.	13
Figure 3.	14
Figure 4.	15
Figure 5.	15
Figure 6.	15
Figure 7.	16
Figure 8.	16
DISCUSSION	17
AUTHORS' CONCLUSIONS	18
ACKNOWLEDGEMENTS	19
REFERENCES	20
CHARACTERISTICS OF STUDIES	24
DATA	40
Test 2. ALFI-MMSE	41
Test 3. IMCT	41
Test 4. SPSMQ	41
Test 5. Tele-Free-Cog	41
Test 7. RUDAS	42
ADDITIONAL TABLES	42
APPENDICES	54
HISTORY	65
CONTRIBUTIONS OF AUTHORS	65
DECLARATIONS OF INTEREST	65
SOURCES OF SUPPORT	65
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	65
INDEX TERMS	66

[Diagnostic Test Accuracy Review]

Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia

Lucy C Beishon¹, Emma Elliott², Tuuli M Hietamies^{2,3}, Riona Mc Ardle⁴, Aoife O'Mahony⁵, Amy R Elliott^{1,6}, Terry J Quinn²

¹Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. ²Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. ³Anesthesiology, Perioperative and Pain Medicine, Stanford University, Stanford, California, USA. ⁴Translational and Clinical Research Institute, Newcastle University, Newcastle, UK. ⁵CUBRIC, School of Psychology, Cardiff University, Cardiff, UK. ⁶NIHR Academic Clinical Fellow, University Hospitals Leicester, Leicester, UK

Contact: Lucy C Beishon, lb330@le.ac.uk.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.

Publication status and date: New, published in Issue 4, 2022.

Citation: Beishon LC, Elliott E, Hietamies TM, Mc Ardle R, O'Mahony A, Elliott AR, Quinn TJ. Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia. *Cochrane Database of Systematic Reviews* 2022, Issue 4. Art. No.: CD013724. DOI: [10.1002/14651858.CD013724.pub2](https://doi.org/10.1002/14651858.CD013724.pub2).

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Remote cognitive assessments are increasingly needed to assist in the detection of cognitive disorders, but the diagnostic accuracy of telephone- and video-based cognitive screening remains unclear.

Objectives

To assess the test accuracy of any multidomain cognitive test delivered remotely for the diagnosis of any form of dementia.

To assess for potential differences in cognitive test scoring when using a remote platform, and where a remote screener was compared to the equivalent face-to-face test.

Search methods

We searched ALOIS, the Cochrane Dementia and Cognitive Improvement Group Specialized Register, CENTRAL, MEDLINE, Embase, PsycINFO, CINAHL, Web of Science, LILACS, and ClinicalTrials.gov (www.clinicaltrials.gov/) databases on 2 June 2021. We performed forward and backward searching of included citations.

Selection criteria

We included cross-sectional studies, where a remote, multidomain assessment was administered alongside a clinical diagnosis of dementia or equivalent face-to-face test.

Data collection and analysis

Two review authors independently assessed risk of bias and extracted data; a third review author moderated disagreements. Our primary analysis was the accuracy of remote assessments against a clinical diagnosis of dementia. Where data were available, we reported test accuracy as sensitivity and specificity. We did not perform quantitative meta-analysis as there were too few studies at individual test level.

For those studies comparing remote versus in-person use of an equivalent screening test, if data allowed, we described correlations, reliability, differences in scores and the proportion classified as having cognitive impairment for each test.

Main results

The review contains 31 studies (19 differing tests, 3075 participants), of which seven studies (six telephone, one video call, 756 participants) were relevant to our primary objective of describing test accuracy against a clinical diagnosis of dementia. All studies were at unclear or high risk of bias in at least one domain, but were low risk in applicability to the review question. Overall, sensitivity of remote tools varied with values between 26% and 100%, and specificity between 65% and 100%, with no clearly superior test.

Across the 24 papers comparing equivalent remote and in-person tests (14 telephone, 10 video call), agreement between tests was good, but rarely perfect (correlation coefficient range: 0.48 to 0.98).

Authors' conclusions

Despite the common and increasing use of remote cognitive assessment, supporting evidence on test accuracy is limited. Available data do not allow us to suggest a preferred test. Remote testing is complex, and this is reflected in the heterogeneity seen in tests used, their application, and their analysis. More research is needed to describe accuracy of contemporary approaches to remote cognitive assessment. While data comparing remote and in-person use of a test were reassuring, thresholds and scoring rules derived from in-person testing may not be applicable when the equivalent test is adapted for remote use.

PLAIN LANGUAGE SUMMARY

How accurate are remote, virtual assessments at diagnosing dementia?

Why is this question important?

Dementia is a chronic and progressive condition that affects people's memory and ability to function day-to-day. A clinical diagnosis of dementia usually involves brain scans, physical examinations and history taking. As a first step, we often use memory and thinking tests to identify people who need further assessment. Traditionally these tests are performed in-person, but modifications of the tests allow them to be used over the telephone or via video calls – sometimes called 'remote assessment'.

The need for remote assessment has become particularly urgent due to COVID-19. However, there are potential benefits of remote assessment beyond the COVID-19 pandemic. Physically attending appointments can be difficult for some people and remote assessments offer greater convenience. Remote assessments are also useful in research, as a large number of people can be reached in a fairly short amount of time.

A test delivered by telephone may not be as good as the in-person equivalent, and getting these tests right is important. On the one hand, if a test suggests someone has dementia when they do not (called a false positive), this can have an emotional impact on the person and their family. On the other hand, not identifying memory and thinking problems when they are present (called a false negative), mean that the person does not get the treatment and support that they need.

What was the aim of this review?

We aimed to assess whether memory and thinking tests carried out by telephone or video call can detect dementia.

What was studied in this review?

We looked at various memory and thinking tests. Many tests have been developed over time and they differ in their content and application, but most are based on a modification of a traditional in-person test.

What were the main results of this review?

The review included 31 studies, using 19 different memory tests, with a total of 3075 participants.

Only seven tests were relevant to our question regarding accuracy of remote testing. With the limited number of studies, estimates on the accuracy of these tests are imprecise. Our review suggests that remote tests could correctly identify people with dementia between 26% and 100% of the time, and could correctly rule out dementia 65% to 100% of the time.

The remaining 24 studies compared a remote test with the face-to-face equivalent. These studies suggested that remote test scores usually agreed with in-person testing, but this was not perfect.

How reliable are the results of the studies in this review?

In these studies, a clinical diagnosis of dementia was used as the reference (gold) standard. We identified a number of issues in the design, conduct and reporting of the studies. A particular issue was around the selection of participants for the studies. Studies often did not include people with hearing or language impairments that may have complicated remote testing.

Who do the results of this study apply to?

Most studies investigated older adults (over 65 years). The findings may not be representative of all older adults with dementia, as some studies only examined specific groups of people, for example, after stroke. The studies were usually performed in specialist centres by experts. So, we do not know how well these tests identify dementia in routine community practice.

What are the implications of this review?

The review highlights the lack of high-quality research describing accuracy of telephone- and video call-based memory and thinking tests. There were many differences between the studies included in this review such as the type of test used, participants included, the setting in which the study is carried out and language studied. This made comparisons between studies difficult. Our review suggests that remote assessments and in-person assessments are not always equivalent. In situations where access to in-person assessment is difficult, remote testing could be used as a useful first step. Ideally, this should be followed up with an in-person assessment before a diagnosis is made. Due to limited studies, and differences in the way studies were carried out, we cannot recommend one particular remote test for the assessment of dementia.

How up to date is this review?

This search was performed in June 2021.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings table test accuracy studies

Patient population	Any adult aged > 18 years requiring cognitive testing.
Index test	Any remote multidomain cognitive screening tool (e.g. telephone, video call, smartphone).
Reference standard	Any clinical diagnosis of dementia.
Target condition	Dementia (all-cause and subtypes).
Included studies	7 studies, 6 studies of telephone assessment (756 participants), 1 study of video call (42 participants).
Quality concerns	All studies were at high or unclear risk of bias for ≥ 1 domain.
Heterogeneity	There was significant heterogeneity between studies, particularly in the populations included, index test used and the reference standard dementia ascertainment.

Study ID	Comparison	Index test	Total number included	Number with dementia	Test threshold	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Telephone assessments									
Burns 2020	Dementia vs no dementia	Tele-Free-Cog	107	64	19	87	100	100	84
					20	90	83	89	85
					21	94	65	80	88
Cammozzato 2011	Alzheimer's disease vs no dementia	ALFI-MMSE (Brazilian)	133	66	13	90	100	100	90
					14	90	89	90	89
					15	94	84	85	93
					16	100	75	80	100

Roccaforte 1992	Dementia (CDR 1–2) vs no dementia (CDR 0–0.5)	ALFI-MMSE	100	81	17	67	100	100	61
Roccaforte 1994	Dementia (CDR 1–2) vs no dementia (CDR 0–0.5)	SPSMQ	100	66	5 – unadjusted	26	97	94	40
					7 – unadjusted	56	91	92	52
					5 – adjusted	41	97	96	46
					7 – adjusted	74	79	87	61
Salazar 2014	N/A	TICS-m MIS-t TEXAS	184	57	NR	NR	NR	NR	NR
Zhou 2004	Dementia vs no dementia	IMCT	132	65	Education adjusted: illiterate – 17 primary school – 20 middle school – 22 university – 23	80	80.6	79	81
Video-call assessments									
Wong 2012	Dementia (moderate or severe impairment) vs no dementia	RUDAS	42	8	23	80	91	73	94

Conclusions

We identified 7 cross-sectional studies examining a remote cognitive screening. We could not perform meta-analysis due to small numbers of studies at test level.

Sensitivity of remote assessment tools varied between 26% and 100% and specificity between 58% and 100%.

Implications

Overall, there is insufficient evidence to recommend the use of any single remote cognitive assessment tool.

There were few eligible studies of even the most commonly used tests and a lack of standardisation in the choice and application of tests.

ALFI-MMSE: Adult Lifestyles and Function Interview Mini-Mental State Examination; CDR: Clinical Dementia Rating; IMCT: Information Memory Concentration Test; MIS-t: Memory Impairment Screen-Telephone; NR: not reported; RUDAS: Rowland Universal Dementia Assessment Scale; SPSMQ: Short Portable Mental Status Questionnaire; Tele-Free-Cog: Telephone Free-Cog; TEXAS: Telephone version of the Executive Interview; TICS-m: Modified Telephone Interview for Cognitive Status.

BACKGROUND

Assessment of cognition using a multidomain test (assessments of differing aspects of cognition) can serve many important purposes (Lin 2013). In clinical practice, cognitive testing may form part of the assessment of the person with a suspected cognitive syndrome, or the testing may be used as an initial triage tool to identify those who need more specialist input. In research, cognitive testing may be used to identify potential participants for a study or as an assessment of treatment effect.

There are many multidomain cognitive assessment tools available to the clinician (Harrison 2016). Indeed, in some areas there are almost as many assessment tools as there are research studies (Lees 2012). An important factor to consider when choosing a cognitive test is the test's accuracy for the detection of the condition of interest, for example the accuracy of a screening test for detection of dementia. In the Cochrane Dementia and Cognitive Improvement Group (CDCIG), we have reviewed the literature and summarised the accuracy of many of the commonly used cognitive screening tests, including Folstein's Mini-Mental State Examination (MMSE) (Creavin 2016), Montreal Cognitive Assessment (MoCA) (Davis 2015), and Addenbrooke's Cognitive Examination (ACE) (Beishon 2019).

To date, our suite of diagnostic test accuracy (DTA) reviews have been limited to in-person, face-to-face assessment, as this is favoured for a clinical diagnosis of dementia and would be usual practice in most services (Davis 2013). The ongoing coronavirus pandemic has caused a fundamental change in practice that no-one had anticipated. The emergency restrictions on movement and social contact necessitated by the viral pandemic limit the opportunity for in-person assessment. Clinical services and research teams have responded, and increasingly consultations are being performed remotely. Various cognitive screening tests designed for administration by telephone or via video call are available and could be well suited to the current situation (Elliott 2020). However, the diagnostic accuracy of these tools should not be assumed, and we considered it was necessary to collate, appraise and present estimates of accuracy for papers describing the use of remote cognitive testing. Remote testing can be performed using the telephone, but increasing availability of audio-visual technologies also allows for assessment using video-based calls or other tele-health approaches. In this review, we were interested in both telephone and video-based assessments.

Beyond the pandemic situation, there are other circumstances in which remotely administered cognitive tests could be useful. Many practitioners in remote and rural areas are already familiar with testing at distance (Barth 2018), and such tests are a core part of a telemedicine memory service (McCleery 2021). In the research context, remote tests may be the only feasible way to include cognitive outcome measures at scale in large pragmatic trials or observational studies (Ritchie 2015).

Target condition being diagnosed

The condition of interest for this review was clinical dementia. We recognise that cognitive testing may be used to inform the diagnosis of other cognitive syndromes such as mild cognitive impairment (MCI), but the condition of greatest relevance is invariably clinical dementia. The dementia diagnosis is operationalised in various classification systems, such as

the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) and the World Health Organization's International Classification of Diseases (ICD) (APA 2013; WHO 2010). Although there are some differences between these classifications (e.g. the most recent DSM guidance suggests use of the terms 'major' and 'mild neurocognitive disorder' rather than dementia and MCI), they all describe dementia as a progressive, irreversible condition characterised by impairments in multiple cognitive domains sufficient to cause problems in activities of daily living. Additional classifications exist to describe pathological dementia subtypes, for example Alzheimer's disease or vascular dementia.

Dementia test accuracy studies have traditionally used the paradigm of assessing a test of interest against a clinical diagnosis of dementia (Takwoingi 2018). For this review, with its focus on remote testing, we anticipated an alternative but equally important study design – where the remote test was compared against the usual, in-person administration of the same test.

Index test(s)

Our index test of interest was any multidomain cognitive assessment tool that was administered remotely, for example by telephone or via video call. We considered these two technologies separately.

We suspected that in most instances the test would be a variation of standard face-to-face cognitive assessment, although content may need to be modified for remote delivery. We were interested in real-time assessment that involved someone administering and scoring the test. Our remit thus did not extend to computer-based cognitive gaming that purports to offer an assessment, self-completed cognitive questionnaires or assessments that exclusively involved a collateral source, for example a family member. The properties of self-complete cognitive questionnaires and informant questionnaires are covered in other Cochrane Reviews (Burton 2021; Hendry 2019).

Several commonly used cognitive screening tools have been adapted for use remotely. The majority of tests that have been adapted from standard in-person assessments (e.g. MMSE, MoCA), removing those components that cannot easily be conducted remotely (e.g. visuospatial tasks), reducing the total score in line with these adaptations. In some instances, these elements are replaced with other cognitive tasks to enhance the discriminative ability of the test. Two common examples include the Modified Telephone Interview for Cognitive Status (TICS-m), and the Adult Lifestyles and Function Interview Mini-Mental State Examination (ALFI-MMSE), both based on the widely used MMSE test. The TICS-m was originally developed as a brief, cognitive assessment tool that could be delivered by telephone for situations where it is not practical to conduct a face-to-face assessment (Brandt 1988; Cook 2009). The TICS-m shares similar test items to MMSE, but was modified to include measures of immediate and delayed recall to improve the sensitivity for dementia detection (Cook 2009; Duff 2015). Similarly, the ALFI-MMSE was developed as a telephone-based cognitive assessment as part of a cohort study which undertook longitudinal assessments on community dwelling adults aged 65 years and older (Roccaforte 1992). The ALFI-MMSE includes similar components to the MMSE, but excludes certain tests of language, motor or visuospatial function, which cannot be conducted by telephone (Roccaforte 1992). Therefore, the ALFI-

MMSE has a total score of 22, although a 26-point version has also been developed (Cammozzato 2011; Roccaforte 1992). In this review, we also included several less widely used remote assessment tools including: memory impairment screen-telephone (MIS-t), telephone version of the executive interview (TEXAS), Information Memory Concentration Test (IMCT), Short Portable Mental Status Questionnaire (SPSMQ), Telephone Free-Cog (Tele-Free-Cog), and the Rowland Universal Dementia Assessment Scale (RUDAS).

Clinical pathway

For consistency with other Cochrane Reviews and in keeping with usual methods, we use the phrase 'diagnostic test accuracy' (Davis 2013). However, we recognise that our index tests of interest are screening in nature and not sufficient to make a clinical diagnosis of dementia on their own. In clinical practice the multidomain test is often the first step in a detailed, multidisciplinary assessment that may also be informed by assessments of function, collateral history, and radiological and laboratory testing (Noel-Storr 2012).

The remote tests of interest are not exclusive to a particular healthcare setting. Brief cognitive screening may be performed in primary care to inform the need for onward referral to specialist services. In secondary care clinic settings, cognitive screening may be performed as part of a diagnostic work-up, or to monitor disease progression. For secondary care services, cognitive screening often forms part of the initial assessment. Many countries recommend early cognitive screening of certain groups such as unscheduled older adult admissions or stroke survivors (Robinson 2015). All of these test scenarios could plausibly be performed remotely, and indeed the current viral pandemic is mandating this approach to testing. In this situation, the purpose of testing is to identify a cognitive baseline and assess for the syndrome of delirium. Prevalence of dementia will vary by setting, and if data allowed, we planned to explore this as part of our investigation of heterogeneity.

Alternative test(s)

Another alternative to in-person cognitive assessment is a questionnaire-based test (Harrison 2015), delivered either by post or using online platforms. We did not consider questionnaire-based approaches in this review, because they were not considered to be directly comparable with remote tests delivered using a trained facilitator.

Rationale

The arguments for timely diagnosis of dementia have been made by various professional societies and will not be rehearsed again here (Robinson 2015). Suffice it to say, cognitive testing is fundamental to the assessment of the person with a suspect cognitive problem. Our motivation for this review was in response to the current global viral pandemic. Clinical and research teams are having to rapidly adapt to new ways of working, and we hoped to provide an evidence base to guide choice of testing. However, for many reasons, it seems likely that remote assessment will remain a feature of healthcare beyond the pandemic.

OBJECTIVES

To assess the test accuracy of any multidomain cognitive test delivered remotely for the diagnosis of any form of dementia.

Secondary objectives

- To describe agreement between a remotely delivered cognitive test and the same or a closely related test delivered in-person.
- To identify the quality and quantity of the research evidence describing test accuracy of remote testing.
- To identify sources of heterogeneity in the test accuracy described.
- To identify gaps in the evidence where further research is required.

Investigation of sources of heterogeneity

Heterogeneity is often seen in clinical test accuracy reviews (Deeks 2001). Important potential sources of heterogeneity included the case-mix of the population being assessed; clinical setting; person performing the assessment; platform used to administer the test (e.g. standard telephone versus video call); threshold scores used to define test positivity and the quality of the included papers. We collected data on all of these factors, and if data allowed, we planned to explore their effect using subgroup and sensitivity analyses as appropriate.

METHODS

Criteria for considering studies for this review

Types of studies

This review has a prepublished protocol (Quinn 2020). Our primary interest was cross-sectional studies, where the index test(s) were administered alongside a clinical diagnosis of dementia.

As a secondary, exploratory analysis we also proposed a review of those papers where a remote assessment was compared to the equivalent in-person test. These papers may or may not also have had information on a clinical diagnosis of dementia. Where data were available that were limited to comparison of remote and in-person testing, we included these studies but the proposed analysis differed, recognising that these studies were assessing correlation/agreement rather than diagnostics. Although not traditional test accuracy studies, we feel these studies provide useful information on how remote tests compare to the original in-person test from which they were derived. This provides useful information on how well the constructs of in-person tests are represented in remote versions of the same tool.

We excluded case-control studies due to the inherent risk of bias and inability to use these data to make any inferences about population predictive value. For the same reasons, we excluded studies that used an enriched sample. For example, studies that only included participants who had a certain screening test score, or where a clinical diagnosis of dementia was limited to participants with a particular cognitive profile.

Although not part of the initial protocol (Quinn 2020), we found papers where the same assessor performed both telephone and in-person equivalent or clinical diagnosis of dementia as part of the same testing session or with minimal time between tests. We considered that these tests were not true assessments of accuracy, would be subject to significant bias, and were more reflective of intra-observer variability of the test and so we excluded them. For a number of studies comparing a remote test to the in-person equivalent, it was unclear whether the same or different assessor

(i.e. independently conducted) conducted the tests. Given this was a secondary objective of the review, we included these studies providing the tests were conducted on different days.

Studies where the index test was compared against future development of a cognitive syndrome (delayed-verification studies) require a differing review approach compared to the traditional cross-sectional test accuracy study. We did not consider delayed-verification studies in this review. We did not include any study where the index and clinical diagnosis of dementia were administered with more than one month between them, as such studies could be considered prognostic rather than diagnostic. Studies where the delay between index and reference was shorter were potentially eligible, and we considered the effect of the delay as part of the risk of bias assessment.

We excluded studies with a small number of cases (fewer than 10), as these studies were unlikely to meaningfully add to our understanding of test accuracy and were prone to various selection biases.

Participants

Our population of interest was any adult (aged over 18 years) requiring cognitive testing.

We did not include studies exclusively comprised of cognitively normal participants, where the purpose of study was to create normative values.

We did not exclude papers on the basis of a selected population, but where the population were not predominantly an older adult group (e.g. studies in traumatic brain injury or in stroke), we noted this and where possible we attempted to explore case-mix as a source of heterogeneity.

We included studies conducted in any healthcare setting, and explored setting as a source of heterogeneity where possible.

Index tests

Studies must have included, not necessarily exclusively, a remote cognitive assessment.

Remote testing involves real-time assessment by a tester (trained or untrained) in a different location to the person being tested. Any platform that allowed remote testing was included, and we anticipated studies using traditional telephone, smartphone and videoconferencing.

Assessments had to be multidomain, as these are the tests used in clinical practice. Tests of single cognitive domains such as attention only were not included.

Included tests may have been modifications of existing in-person tests or bespoke assessments designed for remote use.

The assessment should have been performed remotely, so we did not include studies where, for example, the script of a telephone interview was used in a face-to-face assessment.

Included tests needed to have directly assessed the person of interest. We thus did not include informant-based tests such as AD-8 (Hendry 2019) or Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (Harrison 2015). If a test included

both informant responses and direct testing, and these data were available separately, we included the direct test data. Where tests had contingent scoring, for example if a person scored above a certain value, then further testing was performed, we assessed suitability for inclusion case by case, but did not plan to pool these data with other screening tests.

Where we were comparing a remotely delivered index test to an in-person equivalent, we included those tests from which the remote assessment was derived. For example, the various iterations of the Telephone Interview for Cognitive Status (TICS) were designed to emulate the MMSE (Brandt 1988), and so we considered MMSE as a suitable reference for comparison.

We did not limit the review to a particular remote test strategy, and anticipated including multiple index tests. If data allowed, we considered performing indirect comparisons of estimates of accuracy of various remote tests, but these analyses would be exploratory rather than definitive (Owen 2018).

Our focus for this review was accuracy of testing. We did not formally assess whether remote testing was feasible, acceptable or suitable for the populations being tested, although some of these factors may be relevant to our internal and external validity assessments.

Target conditions

We considered papers reporting any clinical diagnosis of dementia. Dementia diagnosis may have been undifferentiated, or a particular subtype may have been specified. Classifying dementia by subtype was not required for inclusion, but where available these data were recorded. A clinical diagnosis of dementia was considered to be an assessment conducted by a trained health professional (e.g. clinician, psychologist), based on a combination of one or more of the following: clinical history and examination, laboratory investigations, neuroimaging and application of standard criteria for dementia diagnosis.

Reference standards

Our reference standard was a clinical diagnosis of dementia. Within the clinical diagnosis rubric, we included all-cause (unspecified) dementia, using any recognised diagnostic criteria, for example ICD-10 or DSM-IV. For the purposes of this review, and in keeping with other Cochrane DTA Reviews, we considered structured interview assessments such as Clinical Dementia Rating (CDR) as diagnostic (Davis 2013). Dementia diagnoses may specify a pathological subtype, and we included all dementia subtypes in this review. We did not include the cognitive syndrome of delirium in the review (Hendry 2016), and assumed that in the process of making the clinical dementia diagnosis, any reversible causes of cognitive impairment would have been excluded. We recognise that there is no true 'gold' standard assessment of clinical dementia and there is variability in diagnostic label applied even when using validated classification systems (Cerullo 2021).

Studies that made a postmortem diagnosis or based diagnosis on imaging or other biomarkers without corresponding comprehensive clinical assessment were not eligible.

We did not set any limits in relation to severity or stage of dementia. If data were available from sufficient studies, we planned to explore the severity of dementia as a potential source of heterogeneity.

Search methods for identification of studies

Electronic searches

We searched ALOIS, the CDCIG Specialized Register (which includes both intervention and DTA studies in dementia); Cochrane Central Register of Controlled Trials (CENTRAL; 2021, Issue 5; CRSO); MEDLINE (OvidSP); Embase (OvidSP); Web of Science (Clarivate); PsycINFO (OvidSP); CINAHL (EBSCO); LILACS (Latin American and Caribbean Health Science Information database) (BIREME) and ClinicalTrials.gov (www.clinicaltrials.gov/) databases. Each source was searched from inception to June 2021. See [Appendix 1](#) for search strategies run in databases. We used controlled vocabulary such as MeSH terms and Emtree where appropriate. In the searches developed, we made no attempt to restrict studies on the basis of sampling frame or setting. This approach was intended to maximise sensitivity and allow for inclusion on the basis of population-based sampling to be assessed at screening (see [Selection of studies](#)). We did not use search filters (collections of terms aimed at reducing the number needed to screen) as an overall limiter because those that are published have not proved sufficiently sensitive ([Whiting 2008](#)). We did not apply any language restriction to the electronic searches, using translation services as needed. We did not search the grey literature. We performed forward and backward searching of included citations.

Searching other resources

We did not search any additional resources.

Data collection and analysis

Selection of studies

Following searching, titles from various databases were collated in Covidence software ([Covidence 2020](#)), and duplicates removed. The Cochrane Dementia Information Scientist (CF) performed a 'first pass' review, removing clearly irrelevant titles, then a minimum of two review authors (AE, EE, RM, LB, TH, TQ) independently assessed studies for eligibility.

For consistency with our other DTA titles, we adopted a hierarchical approach to exclusion, first excluding on the basis of index test and reference standard, and then on the basis of study methods (case-control, size), and then on the basis of any other reason.

Where a potentially relevant paper was missing data that were needed for analyses, we contacted the primary author by email twice. If the authors did not respond, or the relevant data were not available, we did not include data from this study and labelled it as 'data not suitable for analysis'. In total, we contacted authors from 11 studies ([Castanho 2015](#); [Christodoulou 2016](#); [Crooks 2007](#); [Crooks 2005](#); [Fong 2009](#); [Kennedy 2014](#); [Lanska 1993](#); [Matrisch 2012](#); [Metitieri 2001](#); [Monteiro 1998](#); [Salazar 2014](#)). Of these, two authors responded to confirm data were not available ([Fong 2009](#); [Lanska 1993](#)), and the remainder did not respond. If the same dataset was presented in more than one paper, we included the primary paper but referred to other papers if they contained relevant additional information.

Where studies were described in abstract form, we contacted the lead author(s) to ask if the full paper was published, where possible. We limited the studies included to those published in peer-reviewed scientific journals.

We detailed the study selection process in a PRISMA flow diagram.

Data extraction and management

Two review authors independently extracted data (in pairs from LB, AE, RM, TH, TQ, AM) from eligible papers onto a bespoke data extraction form. The form described population tested, purpose/setting of testing, test(s) administered, details of person performing testing and details of the clinical diagnosis of dementia. We derived components of the 2 × 2 table and prevalence figures.

Following data extraction, the two review authors compared their findings and discussed and resolved any disagreements, with recourse to a third review author (TQ) as needed.

Where a test assigned a score and accuracy data were given for a variety of threshold scores, we collected all of these data in the first instance.

Assessment of methodological quality

Paired, independent raters blinded to each other's scores assessed risk of bias (internal validity) and generalisability (external validity) using the QUADAS-2 tool for test accuracy studies (www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/).

The QUADAS-2 assessment covers issues relating to patient selection, index test, reference standard (clinical diagnosis of dementia) and participant flow. Each domain is assessed for issues related to risk of bias; the first three domains are also assessed for generalisability concerns. Our group has considerable experience using the QUADAS-2 tool. For previous DTA reviews, we convened a group with expertise in test accuracy and dementia to tailor the QUADAS-2 approach to the field of dementia studies. Through this work we have operationalised scoring rules for item- and domain-level QUADAS-2 assessment and modified the generic QUADAS-2 question stems to cover issues pertinent to cognitive testing ([Appendix 2](#)) ([Davis 2013](#)).

We limited our QUADAS-2 assessment to those studies that used a test accuracy design. We did not use QUADAS-2 to assess studies that only considered telephone and equivalent in-person testing as the tool is not designed for this study method.

Statistical analysis and data synthesis

Our primary analysis of interest was accuracy of the various remote assessments against the dichotomous outcome variable 'dementia/no dementia'. To explore this, we applied the recommended Cochrane framework treating each test separately. For each test, where data allowed, we extracted data to populate a standard 2 × 2 data table of binary test results (above and below threshold score) cross-classified against binary reference standard (clinical diagnosis of dementia).

From this table, we calculated sensitivities and specificities, with 95% confidence intervals (CI), at individual test level for thresholds of interest. Primary thresholds of interest were based on the thresholds proposed in the original paper describing the study, unless there was a consensus agreed upon threshold that is used in practice and differs from the original threshold described. We performed these first analyses with standard Review Manager 5 software ([Review Manager 2020](#)), and had planned to use bespoke test accuracy software to support any meta-analyses ([Freeman](#)

2019). We visualised these data using forest plots of sensitivity and specificity.

There were insufficient studies (fewer than three, [Davis 2013](#)) providing test accuracy data for any one index test/reference standard combination, and so quantitative meta-analysis was not performed. Instead, we reported details of each of the tests that reported test accuracy data along with the range of sensitivity and specificity values calculated as described above, or as reported by the studies included in this review.

For assessment of remote versus in-person testing, if data were available as correlations, mean difference, or reliability or agreement measures, we tabulated these and described them in the narrative of the review. We interpreted these results as the level of agreement between a remote test and the in-person equivalent to provide an assessment of how well the constructs of an in-person test are replicated in the remote version of the tool. If scores were presented, we described scores and the proportion classified as having cognitive impairment for each test. Where possible, we described whether scores using the remote assessment were generally higher or lower than the scores when traditional face-to-face assessments were performed.

Investigations of heterogeneity

We did not quantify statistical heterogeneity, or perform quantitative subgroup analyses for these DTA analyses, instead we reported a descriptive summary of potential areas of heterogeneity that are common to DTA studies in the dementia field.

We reviewed forest plots of sensitivity and specificity to identify outliers.

We assessed the following factors as potential sources for heterogeneity:

- populations tested, performing subgroup analyses if specific disease groups featured in the included papers (e.g. traumatic

brain injury, stroke), or if specific healthcare settings featured (e.g. secondary care clinics, where prevalence of disease will differ to other settings such as primary care);

- technical features of the testing strategy (person performing testing, application of the test, platform used for delivering the test; language of testing);
- clinical criteria used to reach dementia diagnosis (e.g. ICD-10; DSM-IV) and the methodology used to reach dementia diagnosis (e.g. individual assessment; group (consensus) assessment).

Sensitivity analyses

We did not perform sensitivity analyses as there was insufficient data available for quantitative meta-analysis.

Assessment of reporting bias

We did not perform a quantitative assessment of reporting bias. We recognised that debate remains around the most robust approach to assessment of reporting bias in DTA ([Wilson 2015](#)), and there was uncertainty on how to apply standard approaches such as funnel plots ([Annefloor van Enst 2014](#)).

RESULTS

Results of the search

In total, our search identified 6590 studies. After removing 289 duplicates, we screened the abstracts of 6301 studies. We then excluded 5999 as being irrelevant to our review question. We assessed the full texts of 302 papers for eligibility, of which we excluded 279 studies. We identified a further eight studies from screening additional sources (e.g. reference lists). The full list of reasons for exclusion is available within the PRISMA diagram ([Figure 1](#)). Of the studies included in the review, only seven provided data on test accuracy against a diagnosis of dementia, of which only one paper assessed video-based screening. In total, 24 studies compared a remote with in-person equivalent test. Details of included studies are available in [Table 1](#).

Figure 1. Study flow diagram.

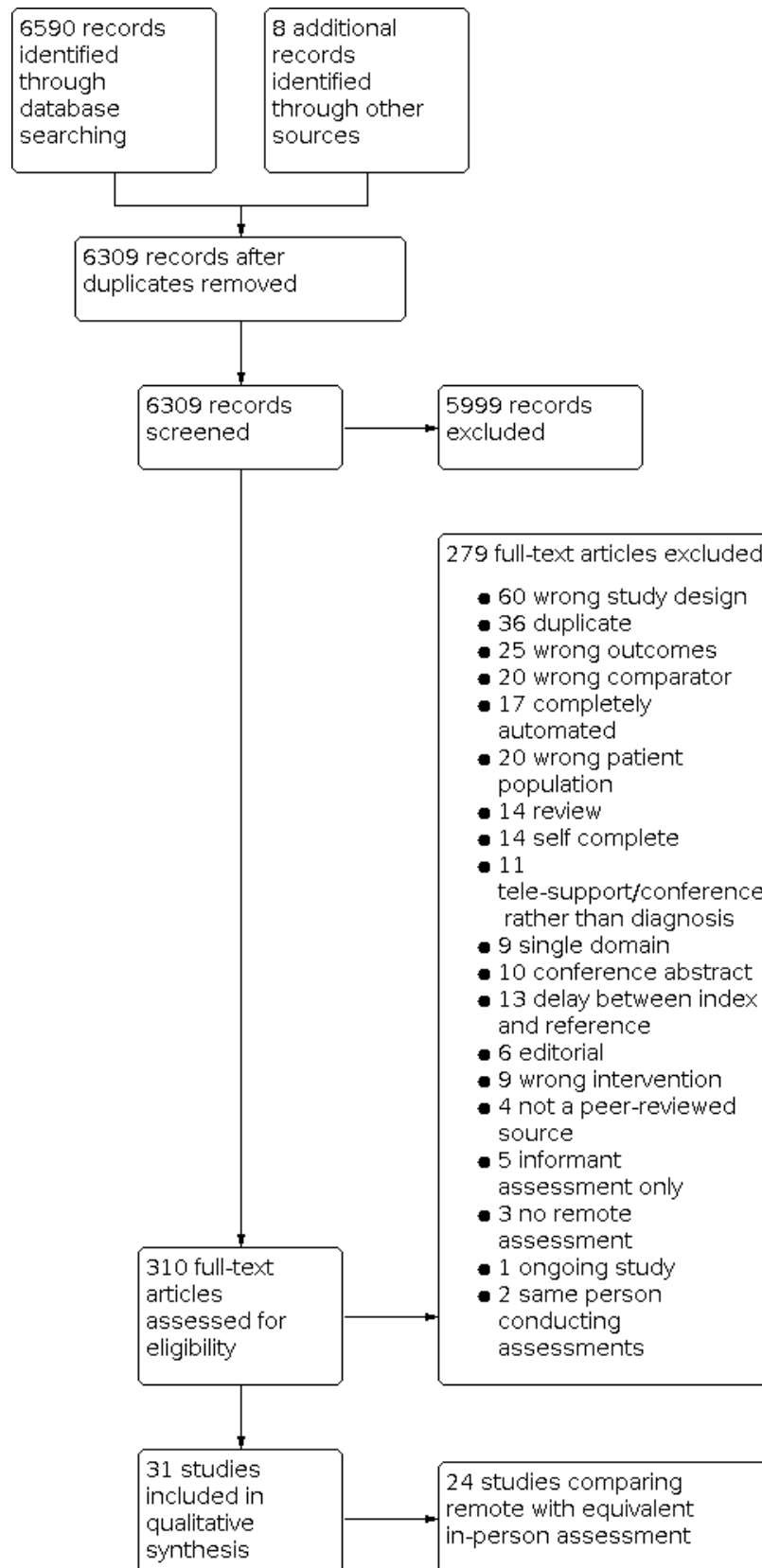
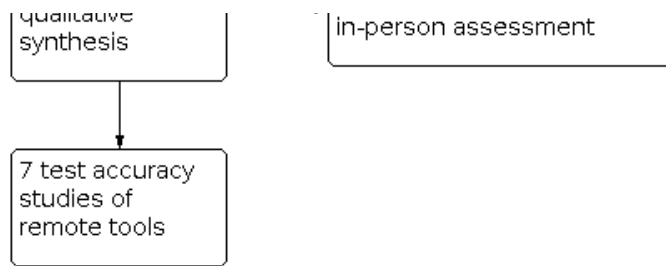


Figure 1. (Continued)



Methodological quality of included studies

We used the QUADAS-2 tool to assess the seven included test accuracy studies for methodological quality and risk of bias. QUADAS-2 results have been presented as graphical displays and as

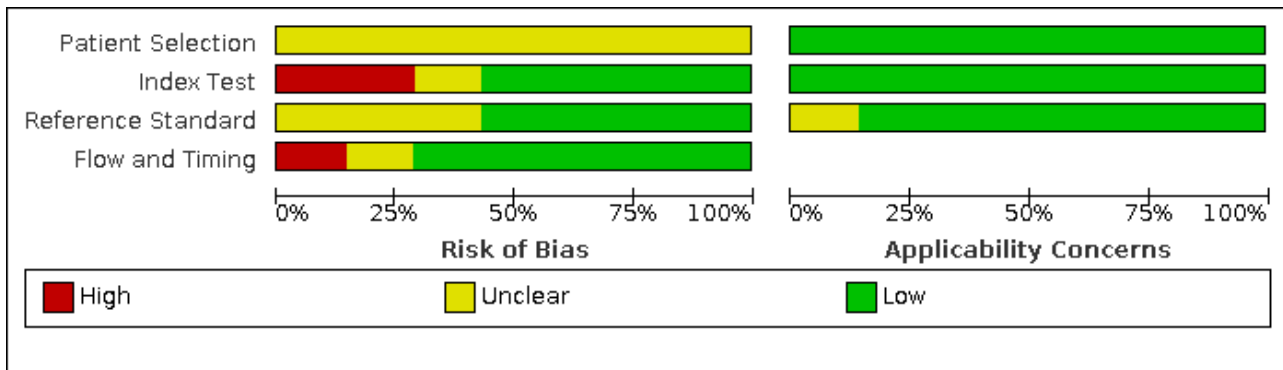
a narrative within the text. There were insufficient data to perform a sensitivity analysis limiting to those papers with low concern for risk of bias across all the QUADAS domains. Full details of the study characteristics and quality assessments can be found in [Figure 2](#); [Figure 3](#).

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

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Burns 2020	?	-	?	?	+	+	?
Cammozzato 2011	?	-	?	-	+	+	+
Roccaforte 1992	?	+	+	+	+	+	+
Roccaforte 1994	?	+	+	+	+	+	+
Salazar 2014	?	+	+	+	+	+	+
Wong 2012	?	?	?	+	+	+	+
Zhou 2004	?	+	+	+	+	+	+

- High	? Unclear	+ Low
---------------	------------------	--------------

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



Patient selection and sampling

We assessed all seven studies at an unclear risk of bias in terms of patient selection and sampling as the sampling strategies used were not reported, or the exclusion criteria were not detailed. Two of these were community-based studies (Cammozzato 2011; Salazar 2014). One study was a combination of outpatient clinics and patients recruited from other research studies (Burns 2020). Two studies were based in outpatient clinics, one in an in-patient rehabilitation unit (Wong 2012), and one was a combination of inpatients and outpatients (Zhou 2004).

In terms of applicability to the review question, we considered all seven studies to be at a low concern.

Index test application

We found that two studies had a high risk of bias in terms of how the index tests were applied (Burns 2020; Cammozzato 2011), and one was unclear (Wong 2012). Neither of the studies assessed at high risk of bias prespecified the test threshold of the remote assessment. The remaining four were found at low risk of bias.

All of the studies were of low concern in terms of the application of the index test and its applicability to the study question.

Reference standard (clinical diagnosis of dementia) application

We found four included studies at low risk of bias in terms of the clinical diagnosis of dementia. The remaining three we assessed as unclear risk (Burns 2020; Cammozzato 2011; Wong 2012). Burns 2020 did not specify sufficient details of the clinical assessment of dementia, including what the gold standard was or how it was assessed. It was unclear from Cammozzato 2011 whether the index test and the clinical diagnosis of dementia were carried out independently from each other.

We had low concerns in terms of the reference standard and its applicability for six studies, there was insufficient information to assess this for Burns 2020.

Flow and timing

When assessing the studies for risk of bias regarding the timing and flow, we found one study was at high risk of bias (Cammozzato 2011), and one unclear (Burns 2020). In Cammozzato 2011,

all participants did not receive the same method of dementia assessment for a clinical diagnosis, and some participants were missing from the analysis. For Burns 2020, it was unclear whether all participants received the same method of assessment for a clinical diagnosis of dementia. The remaining five studies were at low risk of bias.

Findings

We identified six studies that compared a telephone-based assessment of cognitive function to a dementia diagnosis, and one study comparing a video call-based assessment. The details of these studies are summarised in the Characteristics of included studies table.

Of the studies comparing a telephone-based assessment of cognitive function to a dementia diagnosis, one examined the TICS-m (Salazar 2014), two examined the ALFI-MMSE (Cammozzato 2011; Roccaforte 1992), one assessed the MIS-t and the TEXAS (Salazar 2014), one examined the IMCT (Zhou 2004), one examined the SPSMQ (Roccaforte 1994), and one examined the Tele-Free-Cog (Burns 2020).

The study of video-based assessment examined the RUDAS (Wong 2012).

We were unable to perform a quantitative synthesis as too few studies were available for each test.

Twenty-four studies compared a remote assessment to the corresponding in-person test but not a formal dementia diagnosis. Fourteen papers with 15 tests assessed telephone screeners, while 10 described 11 video call-based screening tools. Four studies compared the TICS-m with an in-person MMSE (Baccaro 2015; Castanho 2015; Matrisch 2012) or the similar modified Mini-Mental State Examination (3MSE) (Arnold 2009). One study evaluated the Portuguese version of the TICS-m (Castanho 2015), and one the Brazilian-Portuguese version (Baccaro 2015). Eleven studies examined a telephone or video call version of the MMSE (Ball 1993; Carotenuto 2018; Cullum 2014; Kennedy 2014; Lanska 1993; Loh 2004; Loh 2007; Metitieri 2001; Monteiro 1998; Vahia 2015; Wadsworth 2016), and one examined an Italian version of the telephone MMSE (Itel-MMSE) (Metitieri 2001).

The remaining studies examined different telephone assessments: University of California Screening Battery (UCSF) for amyotrophic

lateral sclerosis (ALS) (Christodoulou 2016), Cognitive Assessment for Later Life Status (CALLS) (Crooks 2007), Ottawa Mental State Exam (OMC) telephone (Dellasega 2001), Telephone Assessment of Mental State (TAMS) (Lanska 1993) and the Brief Cognitive Rating Scale (BCRS) (Monteiro 1998).

Video-based assessments were the Alzheimer's Disease Assessment Cognitive Subscale (ADAS-Cog) (Carotenuto 2018), Short Portable Mental State Examination (SPMSE) (Menon 2001), and RUDAS (Hwang 2022; Wong 2011).

We summarised the findings for each of the included tests comparing against a dementia diagnosis in Summary of findings 1 and those comparing a telephone against an equivalent in-person test in Table 2.

Summary of telephone-based cognitive screening assessments

Modified Telephone Interview for Cognitive Status

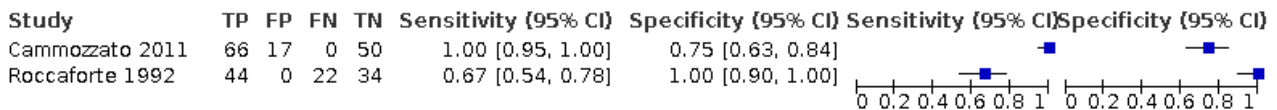
Only one test accuracy study examined the TICS-m but did not report test accuracy data using sensitivity and specificity (Salazar

2014). Higher scores on the TICS-m represent better cognitive function.

Adult Lifestyles and Function Interview Mini-Mental State Examination

Two studies evaluated the ALFI-MMSE (Cammozzato 2011; Roccaforte 1992). Higher scores represent better cognitive function. Cammozato 2011 evaluated a Brazilian translated version of the ALFI-MMSE at four test thresholds: 13 (sensitivity: 90%, specificity: 100%), 14 (sensitivity: 90%, specificity: 89%), 15 (sensitivity: 94%, specificity: 84%), and 16 (sensitivity: 100%, specificity: 75%). Roccaforte 1992 evaluated the ALFI-MMSE at a threshold of 17, relative to the Brief Neuropsychiatric Screening test (BNPS), with a sensitivity of 67% and specificity of 100%. ALFI-MMSE scores correlated with in-person MMSE scores ($r = 0.85$) but this varied with severity of dementia, with a greater correlation with more severe dementia (CDR 0.5: $r = 0.73$, CDR 1: $r = 0.78$, CDR 2: $r = 0.85$) (Roccaforte 1992). Data are summarised in Figure 4 for the ALFI-MMSE at two thresholds.

Figure 4. Forest plot of Adult Lifestyles and Function Interview Mini-Mental State Examination (ALFI-MMSE) at thresholds of 16 and less and 17 and less.

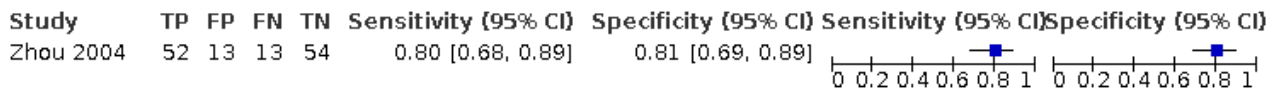


Information Memory Concentration Test

The IMCT is a brief cognitive assessment tool, which has been adapted for remote use (Zhou 2004). The test has a total score of 37 and a higher score is indicative of poorer cognitive function. Zhou

2004 tested a culturally adapted Chinese version of the IMCT, at educationally stratified thresholds (illiterate = 17, primary school = 20, middle school = 22 and university = 23); sensitivity was 80% and specificity was 80.6% for a diagnosis of dementia. Data are summarised in Figure 5.

Figure 5. Forest plot of Information Memory Concentration Test (IMCT) at education adjusted thresholds.

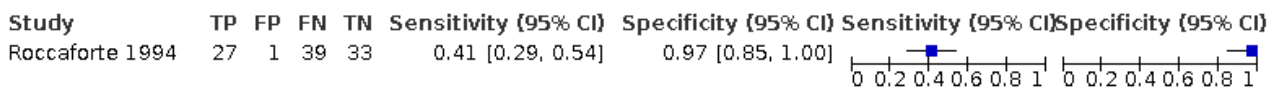


Short Portable Mental Status Questionnaire

The SPMSQ was designed as a brief cognitive assessment tool to test for gross cognitive impairment (Pfeiffer 1975; Roccaforte 1994). The test has a total score of 10 and a threshold of three or four errors identified mild impairment, five to seven identified moderate impairment, and eight more identified severe impairment (Pfeiffer 1975). A threshold of five or greater errors provided the optimal

threshold for identifying dementia (Pfeiffer 1975). At a threshold of five or fewer correct answers, Roccaforte 1994 demonstrated a sensitivity of 41% and specificity of 97% using scores adjusted for ethnicity and education, and a sensitivity of 26% and specificity of 97% for unadjusted scores. At a threshold of seven or greater, sensitivity was 74% (adjusted) and 56% (unadjusted), and specificity was 79% (adjusted) and 91% (unadjusted) (Roccaforte 1994). Data are summarised in Figure 6 for the SPMSQ.

Figure 6. Forest plot of Short Portable Mental Status Questionnaire (SPSMQ) at a threshold of five or less (adjusted score).

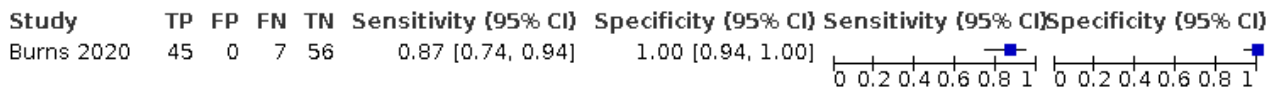


Tele-Free-Cog

Free-Cog is a recently developed cognitive assessment tool designed to be freely available, with a strong focus on the assessment of functional abilities combined with executive function (Burns 2020). The Free-Cog has a total score of 30, where higher scores indicate better cognitive performance. At a test threshold of 26, sensitivity was 83%, and specificity was 80% to

detect dementia (Burns 2020). The Tele-Free-Cog was adapted for use by telephone by removing three components (requiring visuospatial, language or motor skills), giving a total score of 24 (Burns 2020). Burns 2020 determined sensitivity and specificity at three test thresholds: 19 (sensitivity: 87%, specificity: 100%), 20 (sensitivity: 90%, specificity: 83%), 21 (sensitivity: 94%, specificity: 65%). Data are summarised in Figure 7 for the Tele-Free-Cog at a threshold of 19.

Figure 7. Forest plot of Telephone Free-Cog (Tele-Free-Cog) at a threshold of 19 or less.



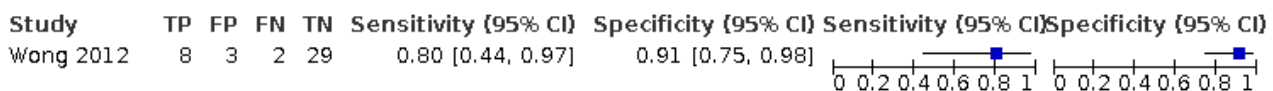
Summary of video call-based cognitive screening assessments

Rowland Universal Dementia Assessment Scale

One study compared a video call-based assessment of the RUDAS (Wong 2012), with a clinical diagnosis of dementia. The RUDAS was

designed for use in culturally and linguistically diverse populations, and higher scores are indicative of better cognitive function (Wong 2012). At a threshold of 23, the RUDAS could identify dementia with a sensitivity of 80%, and a specificity of 91%. Data are summarised in Figure 8 for the RUDAS.

Figure 8. Forest plot of Rowland Universal Dementia Assessment Scale (RUDAS) at a threshold of 23 or less.



Studies comparing a remote versus in-person test

Telephone assessments

Fourteen papers compared a telephone-based test to the in-person equivalent. The method of analysis varied between studies. Six studies compared the TICS-m with the MMSE (Baccaro 2015; Castanho 2015; Matrisch 2012; Plassman 1994), 3MSE (Arnold 2009), or MoCA (Castanho 2015). Compared to the MMSE, the sensitivity was 90% to 92% and specificity was 71% to 74% of the TICS-m at cut-offs of 13.5 and 14 (Baccaro 2015; Castanho 2015). TICS-m scores moderately correlated with MMSE scores (r = 0.48 to 0.66) (Castanho 2015; Matrisch 2012), and strongly with MoCA scores (r = 0.75) (Castanho 2015). Compared to the 3MSE, the sensitivity of the TICS-m to classify low cognition was 80% and specificity was 76% (Arnold 2009). For the MMSE-, MoCA-, 3MSE-, and MMSE-derived remote assessments, higher scores are indicative of better cognitive function.

Six studies investigated a telephone-based version of the MMSE with the in-person test, although not all used the same version of MMSE: ALFI-MMSE (Newkirk 2004), Telephone TAMS (Lanska 1993), Mini-Mental State Examination Telephone (MMSET) (Kennedy 2014), bespoke modified MMSE (Monteiro 1998), telephone MMSE (Garre-Olmo 2008), and Itel-MMSE (Metitieri 2001). Correlations between telephone-based MMSE assessments and the in-person assessment were moderate to strong (r = 0.69 to 0.85) (Kennedy 2014; Lanska 1993; Metitieri 2001), and moderate with the six-item screener (SIS) (r = 0.44) (Kennedy 2014). The MMSET showed good internal consistency with the MMSE (Cronbach a = 0.845 and 0.763) (Kennedy 2014). The TAMS showed strong correlations with the MMSE (r = 0.81), MMSE-TAMS (r = 0.86), ADAS-Cog (r = -0.68), and ADAS (r = -0.80) (Lanska 1993). The intraclass correlation coefficient (ICC) for the telephone and in-person versions of the MMSE was 0.98

(all participants), 0.69 (dementia only) (Monteiro 1998). Monteiro 1998 also compared telephone and in-person versions of the BCRS; ICC: 0.98 and 0.94 (all participants), 0.90 (dementia only).

Christodoulou 2016 compared items from the UCSF by telephone with an in-person MMSE in participants with ALS demonstrating poor to good reliability (ICC range: 0.34 to 0.76). The correlation between the telephone and in-person CALLS was moderate (r = 0.60) (Crooks 2007), and the agreement between OMC scores in-person and by telephone was 100% (Dellasega 2001). Higher scores on the OMC, SPMSQ, and ADAS-Cog indicate poorer cognitive function, whereas lower scores on the UCSF indicate greater impairment.

Video-call assessments

Ten papers compared a video call-based screening test to the in-person equivalent. The majority of studies investigated a video call-based version of the MMSE (Ball 1993; Carotenuto 2018; Cullum 2014; Loh 2007; Loh 2004; Wadsworth 2016; Vahia 2015), two investigated the RUDAS (Hwang 2022; Wong 2011), and one investigated the SPSME (Menon 2001).

The mean test score of the video call and in-person MMSE were comparable in one study (Cullum 2014), higher in the remote test in three studies (Ball 1993; Loh 2007; Vahia 2015), and lower in two studies (Carotenuto 2018; Loh 2004). ICC for the MMSE was 0.89 to 0.90 for in-person versus remote tests (Ball 1993; Loh 2004; Loh 2007).

RUDAS scores were trended to lower in video-call assessments in one study (27 with video call versus 28 with in-person) (Hwang 2022), and were comparable in the other (Wong 2011). In-person and video-call assessments were correlated with a coefficient of

0.73 to 0.79 (Hwang 2022; Wong 2011). Agreement between in-person and remote scores (± 2) was 71% for the RUDAS, and 88% agreement with a clinical diagnosis of dementia (Wong 2011).

One study examined the SPMSE, with a coefficient of variation of 63% with in-person versus 32% with video-call assessments (Menon 2001).

Heterogeneity

Test accuracy studies

The populations enrolled in test accuracy studies included in this review varied from community dwelling older adults (Cammuzzato 2011), participants in research studies (Salazar 2014; Zhou 2004), patients attending an outpatient or geriatric memory service (Burns 2020; Roccaforte 1992; Roccaforte 1994), or hospital inpatients (Wong 2012).

Tests were delivered by a clinical psychologist (Wong 2012), trained or lay interviewer (Cammuzzato 2011), clinician or nurse specialist (Burns 2020; Roccaforte 1992; Roccaforte 1994), bilingual examiner (Salazar 2014), or were not specified (Zhou 2004).

In terms of the reference standard used to determine a clinical diagnosis of dementia, two studies did not report the criteria used for diagnosis (Burns 2020; Wong 2012), one used the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (Cammuzzato 2011), and four used versions of DSM guidance (Roccaforte 1992; Roccaforte 1994; Salazar 2014; Zhou 2004). In addition, two studies used the CDR scale alongside other criteria (Cammuzzato 2011; Roccaforte 1994).

Two studies used a consensus approach to dementia diagnosis (Roccaforte 1994; Salazar 2014). In three studies, diagnosis was by a single clinician (Burns 2020; Wong 2012; Zhou 2004), and two studies did not report the method of dementia assessment (Cammuzzato 2011; Roccaforte 1992).

Studies comparing a remote versus in-person test

The populations enrolled in studies comparing a remote with in-person test included in this review varied from community dwelling older adults (Crooks 2007; Hwang 2022), participants in research studies (Arnold 2009; Baccaro 2015; Kennedy 2014; Monteiro 1998; Newkirk 2004; Plassman 1994), people attending an outpatient or geriatric memory service (Carotenuto 2018; Christodoulou 2016; Cullum 2014; Garre-Olmo 2008; Loh 2007; Metitieri 2001; Vahia 2015; Wadsworth 2016), primary care (Castanho 2015; Dellasega 2001; Matrisch 2012), hospital inpatients (Ball 1993; Loh 2004; Menon 2001; Wong 2011), or not specified (Lanska 1993). Nine studies enrolled specific disease subgroups: Baccaro 2015 enrolled only people with a history of stroke, and Christodoulou 2016 only examined people with a diagnosis of ALS. Carotenuto 2018; Garre-Olmo 2008; Lanska 1993; Loh 2004; Metitieri 2001; Newkirk 2004; and Wadsworth 2016 only included people with a diagnosis of Alzheimer's disease or MCI.

Five studies examined tests in languages other than English: Portuguese (Castanho 2015), Brazilian-Portuguese (Baccaro 2015), Italian (Metitieri 2001), Spanish (Garre-Olmo 2008; Vahia 2015), and one investigated a culturally and linguistically diverse population (Hwang 2022).

The majority of tests were delivered by clinical psychologists (Baccaro 2015; Ball 1993; Carotenuto 2018; Castanho 2015; Cullum 2014; Lanska 1993; Wadsworth 2016), trained or lay interviewers (Christodoulou 2016; Crooks 2007; Plassman 1994; Vahia 2015), clinician/nurse specialist (Ciemens 2009; Dellasega 2001; Hwang 2022; Loh 2004; Loh 2007; Matrisch 2012; Menon 2001; Monteiro 1998; Newkirk 2004; Wong 2011), or were not specified (Arnold 2009; Garre-Olmo 2008; Kennedy 2014; Metitieri 2001). In one study, the assessor was co-located with the participant, and the interpreter delivered the RUDAS by video call (Hwang 2022).

In terms of the in-person test used to compare to a remote assessment, 19 studies used an in-person version of the MMSE as the comparator (Baccaro 2015; Ball 1993; Carotenuto 2018; Castanho 2015; Christodoulou 2016; Crooks 2007; Cullum 2014; Garre-Olmo 2008; Kennedy 2014; Lanska 1993; Loh 2004; Loh 2007; Matrisch 2012; Metitieri 2001; Monteiro 1998; Newkirk 2004; Plassman 1994; Vahia 2015; Wadsworth 2016). Six studies also used in-person versions of the MoCA (Castanho 2015), USCF (Christodoulou 2016), SIS (Kennedy 2014), ADAS (Lanska 1993), ADAS-Cog (Carotenuto 2018), and BCRS (Monteiro 1998) as comparators in addition to the MMSE. One study used the OMC only as the in-person as comparator (Dellasega 2001), one used the SPMSE (Menon 2001), and two used the RUDAS (Hwang 2022; Wong 2011).

DISCUSSION

Summary of main results

In this review we examined evidence for the use of remote cognitive screening tools. Most included studies (20) evaluated telephone-based assessment tools, with a smaller number evaluating video call-based assessments (11). However, only seven papers (six telephone, one video call) met the criteria for our primary test accuracy analysis. The remainder were relevant to our secondary objective of comparing remote tests with the in-person equivalent. We did not identify any studies using smartphone or tablet-based methods of remote assessment.

Given the frequency with which cognitive assessments are performed remotely, both in clinical practice and research, the supporting evidence was surprisingly limited. We found no eligible studies describing diagnostic accuracy of commonly used assessments such as telephone Montreal Cognitive Assessment (tMoCA), the ACE or mini-ACE and others. This may reflect that longer duration assessments (e.g. ACE-III), or those with more extensive testing of visuospatial and executive function, are challenging to deliver remotely. Many of the studies included in this review were conducted prior to the introduction of copyright and training fees for tools such as the MoCA and MMSE, which may explain why fewer historical studies have examined other, now freely available alternatives (e.g. ACE-III).

The majority of studies included in this review assessed a telephone-based version of the MMSE (e.g. TICS-m, ALFI-MMSE). However, even when the remote test was derived from a single parent test, there was substantial heterogeneity in the nomenclature, application, content and scoring of the remote tests. Modifications to tests to allow for remote assessment were not always transparent.

We found risk of bias to be unclear or high for at least one item in all the included studies that contributed to our primary test accuracy analysis. There was no particular pattern of risk of bias across the studies. However, the majority of studies were low risk in terms of applicability to the review question, that is, we considered the tests were applied as they would be in practice.

This limited and inconsistent evidence, with issues around bias, precludes any recommendation on the preferred remote test version, items or test threshold that should be applied in practice. This does not imply that remote assessment is not a valid approach. Taking TICS-m as an example, the sensitivity and specificity reported was broadly comparable with test accuracy metrics described for traditional in-person cognitive assessment tools for the diagnosis of dementia (Beishon 2019; Davis 2015; Harrison 2016).

For those studies that compared remote and equivalent face-to-face testing, there was again substantial heterogeneity. Heterogeneity was particularly evident in the analyses used to compare tests, with correlation, ICC, area under the curve, percentage agreement and other approaches all used. Several studies reported that the correlations between telephone and in-person versions were moderate, with some reporting poor reliability. This is surprising as one would expect test results to be broadly similar if the only difference is the platform used for assessment. These results question how well certain tests adapted for remote use truly capture the constructs of the original in-person test on which they were based. However, to put the results in context, in the original paper by Folstein 1975 describing the MMSE, the test had a test-retest reliability of 0.89, suggesting that even in-person assessments are not always completely reproducible. It is unclear what an 'acceptable' level of correlation is for a remote test to be considered valid (Harvey 2012). All of this is a reminder that thresholds derived for in-person tests may not be suitable for remote equivalents.

Strengths and weaknesses of the review

Strengths of this review include a comprehensive search strategy which was conducted by Information Specialists at the CDCIG. We conducted the review in line with a prepublished protocol, as recommended in guidelines for the conduct of test accuracy reviews of cognitive assessment tools (Davis 2013). We followed best practice in our independent screening, risk of bias assessment and data extraction of included studies.

This review is limited by the small number of eligible studies identified. It could be argued that our inclusion and exclusion criteria were overly stringent. Other reviews of telephone assessment included more studies in their synthesis. We would argue that our approach was designed to ensure that only the highest quality evidence was included. Cognitive assessment is a fundamental part of healthcare and we should not relax our quality control simply because the existing evidence is weak. Similar reviews of in-person cognitive screening have similarly suffered from a lack of high-quality, eligible studies (Beishon 2019; Davis 2015; Harrison 2016).

Applicability of findings to the review question

The primary objective of this review was to assess the test accuracy of any multidomain cognitive test delivered remotely for the clinical

diagnosis of dementia. All included studies focused on either telephone-based or video-call assessment. Most included studies investigated a remote test with an in-person equivalent, rather than formally assessing test accuracy against a clinical diagnosis of dementia. Thus, much of our included evidence is only partly applicable to the main question of interest. Apart from one study (Burns 2020), all included studies were judged to be of low concern in relation to their applicability to the review question.

The populations included in the test accuracy studies varied considerably. Several studies recruited participants from other research studies. There is the potential for bias here as volunteers for a cohort study may differ from a completely unselected clinical population. However, the included studies seemed to have a reasonable case-mix and we did not consider this method of recruitment a threat to the external validity of our results. Of those studies which recruited referrals for assessment of suspected cognitive impairment, the majority were conducted in specialist outpatient clinics or research centres. Only six studies recruited people from primary care or community settings. At present, remote cognitive assessment is more commonly employed by specialist memory services, so the limited evidence base on other settings is not a major concern but we should be mindful about applying these data to primary care or community services. Seven studies included in this review examined specific populations, for example ALS and post-stroke. The findings from these studies can thus only be applied in these specific populations. Several included studies examined the use of remote tests in rural settings, where access to in-person assessment may be challenging. This seems appropriate as this is a context where remote assessment may be especially useful. A small number of studies evaluated translated tools for non-English-speaking populations. No study specifically examined the effect of language or culture on remote cognitive assessment, which may exacerbate the inherent bias of tools that have not been adapted or translated for different populations.

We did not set time limits on our search of the literature and included studies that could be considered 'historical' rather than representing contemporary practice. This is a greater concern for studies of video-based platforms and some included papers (e.g. Ball 1993) were conducted at a time of basic video-call technology.

AUTHORS' CONCLUSIONS

Implications for practice

This review has identified insufficient evidence, both in terms of quantity and quality, to recommend the use of any single remote cognitive test by either telephone or video call. Of the seven studies that investigated test accuracy, the Tele-Free-Cog provided acceptable sensitivity and specificity (83% to 100% at 20 and 21 thresholds), and is one of the few tools not based on an existing copyrighted tool (e.g. Modified Telephone Interview for Cognitive Status (TICS-m), Adult Lifestyles and Function Interview Mini-Mental State Examination (ALFI-MMSE)), which may facilitate widespread adoption more readily. Use of remote assessment was common in clinical and research practice, and has become more frequent due to the pressures and restrictions of the COVID-19 pandemic and the increasing availability and familiarity with remote communication platforms such as videoconferencing. This review highlights a disconnect between the importance and frequency of remote cognitive assessment and the available evidence.

Limited evidence on test accuracy is not synonymous with evidence of limited accuracy. At individual study level the test accuracy data reported were broadly similar to the accuracy described for many common in-person cognitive screening tools. There are situations where in-person testing presents challenges and our data suggest that remote testing could be used in these situations. However, given the limited evidence, in-person assessment should remain the default where possible.

Our data comparing remote testing to the equivalent face-to-face test suggests that there are potential differences in cognitive test performance dependent on the approach to testing. Clinicians should be careful of directly extrapolating cut-offs and scoring rules derived from in-person testing to the remote test scenario.

Implications for research

Our review highlights the need for more robust research describing the accuracy of remote assessments. In particular the evidence around contemporary video-call platforms and technologies such as smartphones was lacking. Given the increasing shift towards remote-based assessments, and increased familiarity with digital technology for both clinicians and patients since the COVID-19 pandemic, a stronger evidence base supporting the use of remote cognitive testing is needed.

A variety of remote cognitive screening tools are described. We found few or no papers describing the accuracy of many of the tools commonly used in practice. The majority of studies investigated a remote version of the Mini-Mental State Examination (MMSE). The use of MMSE in both clinical practice and research is decreasing due to concerns over copyright. Test accuracy studies comparing other remote assessments, that are copyright free, are urgently needed.

We noted substantial inconsistency in the content, delivery and scoring of remote cognitive screening tests, even with tests that shared the same name, for example the multiple iterations of the Telephone Interview for Cognitive Status (TICS). The adaptation and delivery of remote tools needs to be standardised in order to facilitate consistent assessment, comparisons and pooling of study data. As performance may differ when a remote approach is used, best practice guidance, normative values and thresholds need to be investigated that are specific to remote testing.

ACKNOWLEDGEMENTS

We would like to thank peer reviewers Andrew Larner and Dimity Pond and consumer reviewer Cathie Hofstetter for their comments and feedback on the review.

We thank G Martínez and Z Tieges for their contributions to the protocol, and Jenny McCleery for her contribution to the original idea.

REFERENCES

References to studies included in this review

Burns 2020 {published data only}

Burns A, Harrison JR, Symonds C, Morris J.A novel hybrid scale for the assessment of cognitive and executive function: the Free-Cog. *International Journal of Geriatric Psychiatry* 2020;**36**:566-72.

Camozzato 2011 {published data only}

Camozzato AL, Kochhann R, Godinho C, Costa A, Chaves ML.Validation of a telephone screening test for Alzheimer's disease. *Aging, Neuropsychology, and Cognition: a Journal on Normal and Dysfunctional Development* 2011;**18**:180-94.

Roccaforte 1992 {published data only}

Roccaforte WH, Burke WJ, Bayer BL, Wengel SP.Validation of a Telephone Version of the Mini-Mental-State-Examination. *Journal of the American Geriatrics Society* 1992;**40**:697-702.

Roccaforte 1994 {published data only}

Roccaforte WH, Burke WJ, Bayer BL, Wengel SP.Reliability and validity of the Short Portable Mental Status Questionnaire administered by telephone. *Journal of Geriatric Psychiatry & Neurology* 1994;**7**:33-8.

Salazar 2014 {published data only}

Salazar R, Velez CE, Royall DR.Telephone screening for mild cognitive impairment in Hispanics using the Alzheimer's Questionnaire. *Experimental Aging Research* 2014;**40**:129-39.

Wong 2012 {published data only}

Wong L, Martin-Khan M, Rowland J, Varghese P, Gray LC.The Rowland Universal Dementia Assessment Scale (RUDAS) as a reliable screening tool for dementia when administered via videoconferencing in elderly post-acute hospital patients. *Journal of Telemedicine and Telecare* 2012;**18**:176-9.

Zhou 2004 {published data only}

Zhou J, Zhang X, Mundt JC, Wang L, Meng C, Chu C, et al.A comparison of three dementia screening instruments administered by telephone in China. *Dementia* 2004;**3**:69-81.

References to studies excluded from this review

Abdollahi 2016 {published data only}

Abdollahi A, Bull MT, Darwin KC, Venkataraman V, Grana MJ, Dorsey RE, et al.A feasibility study of conducting the Montreal Cognitive Assessment remotely in individuals with movement disorders. *Health Informatics Journal* 2016;**22**:304-11.

Bentvelzen 2019 {published data only}

Bentvelzen AC, Crawford JD, Theobald A, Maston K, Slavin MJ, Reppermund S, et al.Validation and normative data for the Modified Telephone Interview for Cognitive Status: the Sydney Memory and Ageing Study. *Journal of the American Geriatrics Society* 2019;**67**:2108-15.

Chumbler 1998 {published data only}

Chumbler NR, Zhang M.A telephone screening to classify demented older adults. *Clinical Gerontologist: the Journal of Aging and Mental Health* 1998;**19**:79-84.

Ciemins 2009 {published data only}

Ciemins EL, Holloway B, Coon PJ, McClosky-Armstrong T, Min SJ.Telemedicine and the Mini-Mental State Examination: assessment from a distance. *Telemedicine Journal and E-Health* 2009;**15**:476-8.

Cook 2009 {published data only}

Cook SE, Marsiske M, McCoy KJ, Cook SE, Marsiske M, McCoy KJ.The use of the Modified Telephone Interview for Cognitive Status (TICS-M) in the detection of amnesic mild cognitive impairment. *Journal of Geriatric Psychiatry & Neurology* 2009;**22**:103-9.

Crooks 2005 {published data only}

Crooks VC, Clark L, Petitti DB, Chui H, Chiu V.Validation of multi-stage telephone-based identification of cognitive impairment and dementia. *BMC Neurology* 2005;**5**:8.

Cullum 2006 {published data only}

Cullum CM, Weiner MF, Gehrman HR, Hynan LS.Feasibility of telecognitive assessment in dementia. *Assessment* 2006;**13**:385-90.

Fong 2009 {published data only}

Fong TG, Fearing MA, Jonesa RN, Shia P, Marcantonio ER, Rudolph JL.Telephone interview for cognitive status: creating a crosswalk with the Mini-Mental State Examination. *Alzheimer's and Dementia* 2009;**5**:492-7.

Gatz 2002 {published data only}

Gatz M, Reynolds CA, John R, Johansson B, Mortimer JA, Pedersen NL.Telephone screening to identify potential dementia cases in a population-based sample of older adults. *International Psychogeriatrics* 2002;**14**:273-89.

Go 1997 {published data only}

Go RC, Duke LW, Harrell LE, Cody H, Bassett SS, Folstein MF, et al.Development and validation of a Structured Telephone Interview for Dementia Assessment (STIDA): the NIMH Genetics Initiative. *Journal of Geriatric Psychiatry & Neurology* 1997;**10**:161-7.

Graff-Radford 2006 {published data only}

Graff-Radford NR, Ferman TJ, Lucas JA, Johnson HK, Parfitt FC, Heckman MG, et al.A cost effective method of identifying and recruiting persons over 80 free of dementia or mild cognitive impairment. *Alzheimer Disease and Associated Disorders* 2006;**20**:101-4.

Kempen 2007 {published data only}

Kempen GI, Meier AJ, Bouwens SF, van Deursen J, Verhey FR.The psychometric properties of the Dutch version of the Telephone Interview Cognitive Status (TICS). *Tijdschrift voor Gerontologie en Geriatrie* 2007;**38**:38-45.

Knopman 2010 {published data only}

Knopman DS, Roberts RO, Geda YE, Pankratz V, Christianson TJ, Petersen RC, et al. Validation of the Telephone Interview for Cognitive Status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. *Neuroepidemiology* 2010;**34**:34-42.

Larner 2021 {published data only}

Larner AJ. Cognitive testing in the COVID-19 era: can existing screeners be adapted for telephone use? *Neurodegenerative Disease Management* 2021;**11**(1):77-82.

Lipton 2003 {published data only}

Lipton RB, Katz MJ, Kuslansky G, Sliwinski MJ, Stewart WF, Verghese J, et al. Screening for dementia by telephone using the memory impairment screen. *Journal of the American Geriatrics Society* 2003;**51**:1382-90.

Manly 2011 {published data only}

Manly JJ, Schupf N, Stern Y, Brickman AM, Tang MX, Mayeux R. Telephone-based identification of mild cognitive impairment and dementia in a multicultural cohort. *Archives of Neurology* 2011;**68**:607-14.

Martin-Khan 2012 {published data only}

Martin-Khan M, Flicker L, Wootton R, Loh PK, Edwards H, Varghese P, et al. The diagnostic accuracy of telegeriatrics for the diagnosis of dementia via video conferencing. *Journal of the American Medical Directors Association* 2012;**13**(5):19-24.

Myers 2016 {published data only}

Myers CA, Keller JN, Allen HR, Brouillette RM, Foil H, Davis AB, et al. Reliability and validity of a novel Internet-based battery to assess mood and cognitive function in the elderly. *Journal of Alzheimers Disease* 2016;**54**:1359-64.

Tremont 2016 {published data only}

Tremont G, Papandonatos GD, Kelley P, Bryant K, Galioto R, Ott BR. Prediction of cognitive and functional decline using the Telephone-Administered Minnesota Cognitive Acuity Screen. *Journal of the American Geriatrics Society* 2016;**64**:608-13.

Vercambre 2010 {published data only}

Vercambre MN, Cuvelier H, Gayon YA, Hardy-Leger I, Berr C, Trivalle C, et al. Validation study of a French version of the Modified Telephone Interview for Cognitive Status (F-TICS-m) in elderly women. *International Journal of Geriatric Psychiatry* 2010;**11**:1142-9.

Wadsworth 2018 {published data only}

Wadsworth HE, Dhima K, Womack KB, Hart J, Weiner MF, Hynan LS, et al. Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology* 2018;**33**:1040-5.

Wong 2015 {published data only}

Wong A, Nyenhuis D, Black SE, Law LS, Lo ES, Kwan PW, et al. Montreal Cognitive Assessment 5-minute protocol is a brief, valid, reliable, and feasible cognitive screen for telephone administration. *Stroke* 2015;**46**:1059-64.

Younes 2007 {published data only}

Younes M, Hill J, Quinless J, Kilduff M, Peng B, Cook SD, et al. Internet-based cognitive testing in multiple sclerosis. *Multiple Sclerosis* 2007;**13**:1011-9.

Zhou 2020 {published data only}

Zhou J, Chen X, Han Y, Ma M, Liu X. Diagnostic accuracy of cognitive screening tools under different neuropsychological definitions for poststroke cognitive impairment. *Brain and Behavior* 2020;**10**(8):e01671.

Zietemann 2017 {published data only}

Zietemann V, Kopczak A, Muller C, Wollenweber FA, Dichgans M. Validation of the Telephone Interview of Cognitive Status and Telephone Montreal Cognitive Assessment against detailed cognitive testing and clinical diagnosis of mild cognitive impairment after stroke. *Stroke* 2017;**48**:2952-7.

Additional references
Annefloor van Enst 2014

Annefloor van Enst W, Ochodo E, Scholten RJ, Hoofst L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Medical Research Methodology* 2014;**14**:70.

APA 2013

American Psychiatric Association, DSM-5 Task Force. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5th edition. Vol. **xliv**. Washington (DC): American Psychiatric Association, 2013.

Arnold 2009

Arnold AM, Newman AB, Dermond N, Haan M, Fitzpatrick A. Using telephone and informant assessments to estimate missing Modified Mini-Mental State Exam scores and rates of cognitive decline. *Neuroepidemiology* 2009;**33**:55-65.

Baccaro 2015

Baccaro A, Segre A, Wang YP, Brunoni AR, Santos IS, Lotufo PA, et al. Validation of the Brazilian-Portuguese version of the Modified Telephone Interview for cognitive status among stroke patients. *Geriatrics and Gerontology International* 2015;**15**:1118-26.

Ball 1993

Ball CJ, Scott N, McLaren PM, Watson P. Preliminary evaluation of a Low-Cost Video Conferencing (LCVC) system for remote cognitive testing of adult psychiatric patients. *British Journal of Clinical Psychology* 1993;**32**:303-7.

Barth 2018

Barth J, Nickel F. Diagnosis of cognitive decline and dementia in rural areas – a scoping review. *International Journal of Geriatric Psychiatry* 2018;**33**(3):459-74.

Beishon 2019

Beishon LC, Batterham AP, Quinn TJ, Nelson CP, Panerai RB, Robinson T, et al. Addenbrooke's Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia

and mild cognitive impairment. *Cochrane Database of Systematic Reviews* 2019, Issue 12. Art. No: CD013282. [DOI: [10.1002/14651858.CD013282.pub2](https://doi.org/10.1002/14651858.CD013282.pub2)]

Brandt 1988

Brandt J, Spencer M, Folstein M. The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology* 1988;**1**:111-7.

Burton 2021

Burton 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* 2021, Issue 7. Art. No: CD010771. [DOI: [10.1002/14651858.CD010771.pub3](https://doi.org/10.1002/14651858.CD010771.pub3)]

Carotenuto 2018

Carotenuto A, Rea R, Traini E, Ricci G, Fasanaro AM, Amenta F. Cognitive assessment of patients with Alzheimer's disease by telemedicine: pilot study. *JMIR Mental Health* 2018;**5**:e31.

Castanho 2015

Castanho TC, Portugal-Nunes C, Moreira PS, Amorim L, Palha JA, Sousa N, et al. Applicability of the Telephone Interview for Cognitive Status (Modified) in a community sample with low education level: association with an extensive neuropsychological battery. *International Journal of Geriatric Psychiatry* 2015;**31**:128-36.

Cerullo 2021

Cerullo E, Quinn TJ, McCleery J, Vounzoulaki E, Cooper NJ, Sutton AJ. Interrater agreement in dementia diagnosis: a systematic review and meta-analysis. *International Journal of Geriatric Psychiatry* 2021;**36**(8):1127-47.

Christodoulou 2016

Christodoulou G, Gennings C, Hupf J, Factor-Litvak P, Murphy J, Goetz RR, et al. Telephone based cognitive-behavioral screening for frontotemporal changes in patients with amyotrophic lateral sclerosis (ALS). *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 2016;**17**:1-7.

Ciemens 2009

Ciemens EL, Holloway B, Coon PJ, McClosky-Armstrong T, Min SJ. Telemedicine and the Mini-Mental State Examination: assessment from a distance. *Telemedicine Journal and E-Health* 2009;**15**:476-8.

Covidence 2020 [Computer program]

Veritas Health Innovation Covidence. Version accessed 5 April 2022. Melbourne, Australia: Veritas Health Innovation, 2020. www.covidence.org.

Creavin 2016

Creavin ST, Wisniewski S, Noel-Storr AH, Trevelyan CM, Hampton T, Rayment D, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of*

Systematic Reviews 2016, Issue 1. Art. No: CD011145. [DOI: [10.1002/14651858.CD011145.pub2](https://doi.org/10.1002/14651858.CD011145.pub2)]

Crooks 2007

Crooks VC, Parsons TD, Buckwalter JG. Validation of the Cognitive Assessment of Later Life Status (CALLS) instrument: a computerized telephonic measure. *BMC Neurology* 2007;**7**:10.

Cullum 2014

Cullum CM, Hynan LS, Grosch M, Parikh M, Weiner MF. Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society* 2014;**20**:1028-33.

Davis 2013

Davis DH, Creavin ST, Noel-Storr A, Quinn TJ, Smailagic N, Hyde C, et al. Neuropsychological tests for the diagnosis of Alzheimer's disease dementia and other dementias: a generic protocol for cross-sectional and delayed-verification studies. *Cochrane Database of Systematic Reviews* 2013, Issue 3. Art. No: CD010460. [DOI: [10.1002/14651858.CD010460](https://doi.org/10.1002/14651858.CD010460)]

Davis 2015

Davis DH, Creavin ST, Yip JL, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane Database of Systematic Reviews* 2015, Issue 10. Art. No: CD010775. [DOI: [10.1002/14651858.CD010775.pub2](https://doi.org/10.1002/14651858.CD010775.pub2)]

Deeks 2001

Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**:157-62.

Dellasega 2001

Dellasega CA, Lacko L, Singer H, Salerno F. Telephone screening of older adults using the Orientation-Memory Concentration Test. *Geriatric Nursing* 2001;**22**:253-7.

Duff 2015

Duff K, Tometch D, Dennett K. The Modified Telephone Interview for Cognitive Status is more predictive of memory abilities than the Mini-Mental State Examination. *Journal of Geriatric Psychiatry and Neurology* 2015;**28**:193-7.

Elliott 2020

Elliott E, Green C, Llewellyn DJ, Quinn TJ. Accuracy of telephone-based cognitive screening tests: systematic review and meta-analysis. *Current Alzheimer's Research* 2020;**17**(5):460-71.

Folstein 1975

Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;**12**:189-98.

Freeman 2019

Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: metaDTA. *BMC Medical Research Methodology* 2019;**19**(1):81.

Garre-Olmo 2008

Garre-Olmo J, Lax-Pericall C, Turro-Garriga O, Soler-Cors O, Monserrat-Vila S, Vilalta-Franch J, et al. Adaptation and convergent validity of a telephone-based Mini-Mental State Examination [Adaptación y validez convergente de una versión telefónica del Mini-Mental State Examination]. *Medicina Clínica* 2008;**131**:89-95.

Harrison 2015

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 secondary care setting. *Cochrane Database of Systematic Reviews* 2015, Issue 3. Art. No: CD010772. [DOI: [10.1002/14651858.CD010772.pub2](https://doi.org/10.1002/14651858.CD010772.pub2)]

Harrison 2016

Harrison K, Noel-Storr AH, Demeyere N, Reynish EL, Quinn TJ. Outcomes measures in a decade of dementia and mild cognitive impairment trials. *Alzheimer's Research & Therapy* 2016;**8**(1):48.

Harvey 2012

Harvey PD. Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience* 2012;**14**:91-9.

Hendry 2016

Hendry K, Quinn TJ, Evans J, Scortichini V, Miller H, Burns J, et al. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. *Age and Ageing* 2016;**45**:832-7.

Hendry 2019

Hendry K, Green C, McShane R, Noel-Storr AH, Stott DJ, Anwer S, et al. AD-8 for detection of dementia across a variety of healthcare settings. *Cochrane Database of Systematic Reviews* 2019, Issue 3. Art. No: CD011121. [DOI: [10.1002/14651858.CD011121.pub2](https://doi.org/10.1002/14651858.CD011121.pub2)]

Hwang 2022

Hwang K, De Silva A, Simpson JA, LoGiudice D, Engel L, Gilbert AS, et al. Video-interpreting for cognitive assessments: an intervention study and micro-costing analysis. *Journal of Telemedicine and Telecare* 2022;**28**(1):58-67.

Kennedy 2014

Kennedy RE, Williams CP, Sawyer P, Allman RM, Crowe M. Comparison of in-person and telephone administration of the Mini-Mental State Examination in the University of Alabama at Birmingham Study of Aging. *Journal of the American Geriatrics Society* 2014;**62**:1928-32.

Lanska 1993

Lanska DJ, Schmitt FA, Stewart JM, Howe JN. Telephone-Assessed Mental State. *Dementia* 1993;**4**:117-9.

Lees 2012

Lees R, Fearon P, Harrison JK, Broomfield NM, Quinn TJ. Cognitive and mood assessment in stroke research: focused review of contemporary studies. *Stroke* 2012;**43**(6):1678-80.

Lin 2013

Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: a systematic review for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 2013;**159**(9):601-12.

Loh 2004

Loh PK, Ramesh P, Maher S, Saligari J, Flicker L, Goldswain P. Can patients with dementia be assessed at a distance? The use of Telehealth and standardised assessments. *Internal Medicine Journal* 2004;**34**:239-42.

Loh 2007

Loh PK, Donaldson M, Flicker L, Maher S, Goldswain P. Development of a telemedicine protocol for the diagnosis of Alzheimer's disease. *Journal of Telemedicine and Telecare* 2007;**13**:90-4.

Matrisch 2012

Matrisch M, Trampisch U, Klaaßen-Mielke R, Pientka L, Trampisch HJ, Thiem U. Screening for dementia using telephone interviews. An evaluation and reliability study of the Telephone Interview for Cognitive Status (TICS) in its modified German version [Demenzscreening per Telefon]. *Zeitschrift für Gerontologie und Geriatrie* 2012;**45**:218-23.

McCleery 2021

McCleery J, Lavery J, Quinn TJ. Diagnostic test accuracy of telehealth assessment for dementia and mild cognitive impairment. *Cochrane Database of Systematic Reviews* 2021, Issue 7. Art. No: CD013786. [DOI: [10.1002/14651858.CD013786.pub2](https://doi.org/10.1002/14651858.CD013786.pub2)]

Menon 2001

Menon AS, Kondapavalru P, Krishna P, Chrismer JB, Raskin A, Hebel JR, et al. Evaluation of a portable low cost videophone system in the assessment of depressive symptoms and cognitive function in elderly medically ill Veterans. *Journal of Nervous and Mental Disease* 2001;**189**:399-401.

Metitieri 2001

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

Monteiro 1998

Monteiro IM, Boksay I, Auer SR, Torossian C, Sinaiko E, Reisberg B. Reliability of routine clinical instruments for the assessment of Alzheimer's disease administered by telephone. *Journal of Geriatric Psychiatry and Neurology* 1998;**11**:18-24.

Newkirk 2004

Newkirk LA, Kim JM, Thompson JM, Tinklenberg JR, Yesavage JA, Taylor JL. Validation of a 26-Point telephone version of the Mini-Mental State Examination. *Journal of Geriatric Psychiatry and Neurology* 2004;**17**:81-7.

Noel-Storr 2012

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* 2012;**83**(4):364-73.

Owen 2018

Owen RK, Cooper NJ, Quinn TJ, Lees R, Sutton AJ. Network meta-analysis of diagnostic test accuracy studies identifies and ranks the optimal diagnostic tests and thresholds for health care policy and decision-making. *Journal of Clinical Epidemiology* 2018;**99**:64-74.

Pfeiffer 1975

Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society* 1975;**23**:433-41.

Plassman 1994

Plassman BL, Newman TT, Welsh KA, Helms M, Breitner JC. Properties of the telephone interview for cognitive status: application in epidemiological and longitudinal studies. *Neuropsychiatry, Neuropsychology and Behavioral Neurology* 1994;**7**:235-41.

Review Manager 2020 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.4. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2020.

Ritchie 2015

Ritchie CW, Terrera GM, Quinn TJ. Dementia trials and dementia tribulations: methodological and analytical challenges in dementia research. *Alzheimer's Research & Therapy* 2015;**7**(1):31.

Robinson 2015

Robinson L, Tang E, Taylor JP. Dementia: timely diagnosis and early intervention. *BMJ* 2015;**350**:h3029.

Takwoingi 2018

Takwoingi Y, Quinn TJ. Review of diagnostic test accuracy (DTA) studies in older people. *Age and Ageing* 2018;**47**(3):349-55.

Vahia 2015

Vahia IV, Ng B, Camacho A, Cardenas V, Cherner M, Depp C, et al. Telepsychiatry for neurocognitive testing in older rural Latino adults. *American Journal of Geriatric Psychiatry* 2015;**23**:666-70.

Wadsworth 2016

Wadsworth HE, Galusha-Glasscock JM, Womack KB, Quiceno M, Weiner MF, Hynan LS, et al. Remote neuropsychological assessment in rural American Indians with and without cognitive impairment. *Archives of Clinical Neuropsychology* 2016;**33**:1040-5.

Whiting 2008

Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. *Journal of Clinical Epidemiology* 2008;**41**(4):357-64.

WHO 2010

World Health Organization. International Statistical Classification of Diseases and Related Health Problems (ICD). Vol. 2. Geneva: World Health Organization, 2010.

Wilson 2015

Wilson C, Kerr D, Noel-Storr A, Quinn TJ. Associations with publication and assessing publication bias in dementia diagnostic test accuracy studies. *International Journal of Geriatric Psychiatry* 2015;**30**(12):1250-6.

Wong 2011

Wong L, Martin-Khan M, Rowland J, Varghese P, Gray LC. Reliability of the Rowland Universal Dementia Assessment Scale (RUDAS) via video conferencing. *International Journal of Geriatric Psychiatry* 2011;**26**:988-9.

References to other published versions of this review

Quinn 2020

Quinn TJ, Elliott E, Hietamies TM, Martínez G, Tiegies Z, Mc Ardle R. Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia. *Cochrane Database of Systematic Reviews* 2020, Issue 9. Art. No: CD013724. [DOI: [10.1002/14651858.CD013724](https://doi.org/10.1002/14651858.CD013724)]

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Burns 2020

Study characteristics

Patient Sampling	Participants recruited from non-specialist, outpatient secondary care settings across the UK, where patients were referred for memory assessment (non-specialist memory clinics, community mental health teams), or drawn from other research studies.
	<u>Sampling procedure</u> : not reported.
	<u>Inclusion criteria</u> : people who were referred for a memory assessment or had a diagnosis of dementia/memory impairment as dementia/MCI cases. Any participants with no

Burns 2020 (Continued)

diagnosis of dementia/memory impairment and no self-reported concerns of memory impairment were included as healthy controls.

Exclusion criteria: aged < 18 years

Patient characteristics and setting

107 participants attending outpatient memory services across the UK. Clinical diagnoses were recorded under the following categories: dementia (Alzheimer's disease, frontotemporal dementia, Lewy body dementia, vascular dementia, Parkinson's disease dementia, early-onset dementia, other dementias, mixed Alzheimer/vascular dementia); MCI; or controls (defined as people without any symptoms suggestive of cognitive decline).

Mean age (years): dementia: 77.3 (SD 8.7), control: 63.1 (SD 14.4)

Sex (female): dementia: 44%, control: 70%

Years in education: not reported.

Ethnicity: not reported.

Index tests

Index test: Free-Cog and Tele-Free-Cog

Other tests: MMSE, MoCA, ACE

No specific training was provided and clinicians asked to use judgement. Assessors were either clinicians or experts in conducting cognitive tests.

Test thresholds were not prespecified and optimal thresholds were determined from test data.

It was not clear whether the test was conducted independently of the reference standard.

Target condition and reference standard(s)

Reference standard: normal clinical assessment, including cognitive tests the clinic usually applied.

No specific criteria were used for the diagnosis of dementia. No details were provided on the conduct of the reference standard.

It was unclear whether the reference standard was conducted independently of the index test.

Flow and timing

The Free-Cog was conducted at the same time or within 1 week of the reference standard (unclear whether this also applied to the Tele-Free-Cog).

12 participants were excluded due to diagnosis omission or under remained investigation.

Sensitivity and specificity of the Tele-Free-Cog varied by test threshold (≤ 19 : 87%, 100%, ≤ 20 : 90%, 83%, ≤ 21 : 94%, 65%). True and false positives and negatives were not provided, so data were calculated using Review Manager 5 (Review Manager 2020).

Comparative

Notes

Methodological quality
Item
Authors' judgement
Risk of bias
Applicability concerns
DOMAIN 1: Patient Selection

Burns 2020 (Continued)

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Unclear risk
Are there concerns that the included patients and setting do not match the review question?	Low concern
DOMAIN 2: Index Test (All tests)	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	No
Could the conduct or interpretation of the index test have introduced bias?	High risk
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern
DOMAIN 3: Reference Standard	
Is the reference standards likely to correctly classify the target condition?	Unclear
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear
DOMAIN 4: Flow and Timing	
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Unclear
Were all patients included in the analysis?	No

Burns 2020 (Continued)

Could the patient flow have introduced bias?

Unclear risk

Cammozzato 2011

Study characteristics

Patient Sampling	<p>People aged > 60 years and living within a catchment area of the Hospital de Clinicas de Porto Alegre Brazil were invited to participate in the Alzheimer's Disease center and neurogeriatric clinic at the hospital.</p> <p><u>Sampling procedure:</u> not reported.</p> <p><u>Inclusion criteria:</u> aged > 60 years, in catchment area of the Hospital de Clinicas de Porto Alegre Brazil.</p> <p><u>Exclusion criteria:</u> history of deafness, complaint of hearing impairment, positive whispered voice screening test and MMSE < 11, severe dementia, people with AD with pre-existing psychiatric conditions and with severe clinical comorbidities.</p>
Patient characteristics and setting	<p>Community sample of people with AD and healthy participants in Brazil. Participants underwent a standardised clinical, psychiatric, neuropsychological and neurological evaluation. A collateral informant was used to verify the history. Participants whose CDR score was ≥ 0.5 also underwent neuroimaging and blood tests. Groups were comparable for age, education and gender.</p> <p><u>Total sample:</u> 66 people with AD, 67 controls</p> <p><u>Mean age (years):</u> Group 1: AD: 72.8 (SD 5.6), controls: 70.5 years (SD 6.5); Group 2: AD: 72.6 (SD 7.3), controls: 71.5 (SD 7.0); Group 3: AD: 75.8 (SD 6.4), controls: 74.5 (SD 6.3); Group 4: AD: 74.4 (SD 4.8), controls: 69.2 (SD 8.1)</p> <p><u>Sex (female):</u> Group 1: AD: 47%, controls: 69%; Group 2: AD: 59%, controls: 53%; Group 3: AD: 65%, controls: 77%; Group 4: AD: 47%, controls: 59%</p> <p><u>Mean years in education:</u> Group 1: AD: 3.9 (SD 1.2), controls: 4.8 (SD 1.6); Group 2: AD: 5.4 (SD 4.8), controls: 5.1 (SD 3.6); Group 3: AD: 5.5 (SD 3.6), controls: 5.3 (SD 2.8); Group 4: AD: 5.1 (SD 2.6), controls: 5.9 (SD 3.5)</p> <p><u>Ethnicity:</u> not reported.</p>
Index tests	<p><u>Index test:</u> ALFI-MMSE</p> <p><u>Other tests:</u> none</p> <p>The ALFI-MMSE was translated to Brazilian-Portuguese and back translated to English before being applied. No information on the translation or adaptation procedures. Participants were randomised to 4 groups according to test order: Group 1: ALFI-MMSE followed by ALFI-MMSE administration; Group 2: in-person MMSE followed by in-person MMSE; Group 3: in-person MMSE followed by ALFI-MMSE; Group 4: ALFI-MMSE followed by in-person MMSE.</p> <p>Previously trained interviewers conducted tests, independent of the knowledge of the in-person assessments and diagnoses.</p> <p>Test thresholds were not prespecified and optimal cut-offs were determined using study data.</p>
Target condition and reference standard(s)	<p><u>Target condition:</u> AD</p> <p><u>Reference standard:</u> NINCDS-ADRDA</p>

Cammozato 2011 (Continued)

No information was provided on the reference standard assessors. It was unclear whether the reference standard was blinded to the knowledge of the index tests.

Flow and timing

The time interval between the in-person and telephone assessment was 48–72 hours.

Not all participants received the same reference standard, only participants with CDR \geq 0.5 underwent additional testing (blood and neuroimaging tests).

The sensitivity of the ALFI-MMSE was 94% and specificity was 84% (cut-off of 15). No data were provided on true and false positives and negatives and these were calculated using Review Manager 5 (Review Manager 2020).

No data were provided on dropouts from the study.

Comparative

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Cammozzato 2011 (Continued)

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern
DOMAIN 4: Flow and Timing	
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	High risk

Roccaforte 1992
Study characteristics

Patient Sampling	<p>People attending the University of Nebraska Geriatric Assessment Centre were invited by the research team to take part.</p> <p><u>Sampling procedure:</u> all successive patients who contacted the clinic were invited to enrol. The Geriatric Assessment Centre is an assessment clinic so would be people contacting to self-refer or refer clients of social services or family members/loved ones.</p> <p><u>Inclusion criteria:</u> patients referred or self-referred to Nebraska Geriatric Assessment Centre.</p> <p><u>Exclusion criteria:</u> not stated other than not completing the assessment process or not consenting.</p>
Patient characteristics and setting	<p>175 potential participants. 109 (62%) were willing to participate. 100 completed both the telephone interviews and the geriatric assessment.</p> <p><u>Total number:</u> 100 participants</p> <p><u>Age:</u> 27% aged 65–74 years; 53% aged 75–84; 20% aged ≥ 85 years</p>

Roccaforte 1992 (Continued)

Sex (% female): 76%

Years in education: 22% ≤ 8 years; 51% 9–12 years; 5% > 12 years; 2 people unknown level of education experience

Ethnicity: 95% white, 5% black

Index tests

Index test: ALFI-MMSE

Other tests: collateral version of ALFI-MMSE

ALFI-MMSE conducted by telephone. All conducted by 1 clinical nurse specialist. A collateral version of the ALFI-MMSE was also obtained in a separate telephone call.

Target condition and reference standard(s)

Target condition: dementia

Reference standard: participants undergo a comprehensive in-person assessment including history, examination, blood tests and imaging, as well as other neuropsychological tests, including face-to-face MMSE and BNPS administered by a different nurse specialist to the telephone interview. Diagnoses of dementia made according to DSM III-R criteria, and severity rated according to CDR score.

Flow and timing

Mean length of time between clinic assessment and telephone interview was 8.7 days. Telephone interview was always performed first. 9 participants missing from the analysis.

Comparative

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Roccaforte 1992 (Continued)

Could the conduct or interpretation of the index test have introduced bias?	Low risk
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern
DOMAIN 3: Reference Standard	
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern
DOMAIN 4: Flow and Timing	
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Low risk

Roccaforte 1994
Study characteristics

Patient Sampling	<p>People who contacted the University of Nebraska Geriatric Assessment Centre were invited to take part. Same dataset as the 1992 paper.</p> <p><u>Sampling procedure:</u> all successive patients who contacted the clinic were invited to enrol. Geriatric Assessment Centre is an assessment clinic so would be people contacting to self-refer or refer clients of social services or family members/loved ones.</p> <p><u>Inclusion criteria:</u> patients referred or self-referred to Nebraska Geriatric Assessment Centre.</p> <p><u>Exclusion criteria:</u> not stated other than not completing the assessment process or not consenting.</p>
Patient characteristics and setting	<p>175 potential participants. 109 (62%) were willing to participate. 100 completed both the telephone interviews and the geriatric assessment.</p> <p><u>Total number:</u> 100 participants</p>

Roccaforte 1994 (Continued)

Age: 27% aged 65–74 years; 53% aged 75–84 years; 20% aged ≥ 85 years

Sex (% female): 76%

Year in education: 22% ≤ 8 years; 51% 9–12 years; 25% > 12 years; 2 people unknown level of education experience

Ethnicity: 95% white, 5% black

Index tests	<p><u>Index test</u>: SPMSQ</p> <p><u>Other tests</u>: ALFI-MMSE</p> <p>Test conducted by telephone within the ALFI-MMSE questionnaire. SPMSQ made up of demographic and memory-testing portions of the ALFI-MMSE. Scoring of SPMSQ based on number of correct answers, in contrast to previous studies which have classified according to number of errors made, and adjusted for education and race.</p> <p>In this study, scores were adjusted for race and education. ≤ 7 correct answers provided optimal sensitivity and specificity for diagnosing dementia.</p>		
Target condition and reference standard(s)	<p><u>Target condition</u>: dementia</p> <p><u>Reference standard</u>: in-depth in-person assessment including history, physical examination, laboratory tests and imaging, and neuro-psychological tests.</p>		
Flow and timing	<p>All participants received the index test and reference standard. Telephone assessment occurred first with a mean interval of 8.7 days in between telephone and in-person assessment (SD 6.4; range 3–35 days). 9 participants missing from the analysis.</p>		
Comparative			
Notes			
Methodological quality			
Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Roccaforte 1994 (Continued)

If a threshold was used, was it pre-specified?	Yes	
Could the conduct or interpretation of the index test have introduced bias?		Low risk
Are there concerns that the index test, its conduct, or interpretation differ from the review question?		Low concern
DOMAIN 3: Reference Standard		
Is the reference standards likely to correctly classify the target condition?	Yes	
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		Low risk

Salazar 2014
Study characteristics

Patient Sampling	<p>184 participants enrolled in longitudinal, observational study (Texas Alzheimer's Research and Care Consortium) at the San Antonio site.</p> <p><u>Sampling procedure:</u> not stated.</p> <p><u>Inclusion criteria:</u> not stated.</p> <p><u>Exclusion criteria:</u> not stated.</p>
Patient characteristics and setting	<p>Participants enrolled in a longitudinal, observational cohort study (Texas Alzheimer's Research and Care Consortium), at the San Antonio site.</p> <p><u>Total number:</u> 184</p> <p><u>Mean age (years):</u> 67.65 (SD 7.17)</p>

Salazar 2014 (Continued)

	<p><u>Sex (% female)</u>: 55%</p> <p><u>Mean years in education</u>: 13.6 (SD 2.7)</p> <p><u>Ethnicity</u>: 100% Mexican Chicano/Mexican Americans. All Hispanic</p>
Index tests	<p><u>Index test</u>: AQ – informant based</p> <p><u>Other tests</u>: TICS-m, MIS-T and TEXAS</p>
Target condition and reference standard(s)	<p><u>Target condition</u>: MCI and dementia</p> <p><u>Reference standard</u>: presence or absence of dementia assessed by an in-person clinical interview and neurological examination, and neuropsychological battery assessment. Diagnoses made according to DSM-IV criteria and severity assessed using CDR.</p>
Flow and timing	<p>The interval between the index test and reference standard was not stated. All participants received an index and a reference standard. 2 participants missing from the final analysis.</p>
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Salazar 2014 (Continued)

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern
DOMAIN 4: Flow and Timing	
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Low risk

Wong 2012
Study characteristics

Patient Sampling	<p>Participants were recruited from an in-patient geriatric rehabilitation unit. No additional information provided.</p> <p><u>Sampling procedure</u>: the treating team identified participants based on the inclusion and exclusion criteria.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: severe hearing impairment, need for an interpreter, severe dysphasia, medical instability, delirium.</p>
Patient characteristics and setting	<p>In-patients at a geriatric rehabilitation unit.</p> <p><u>Total number</u>: 42</p> <p><u>Mean age (years)</u>: 75 (range 41–95)</p> <p><u>Sex</u>: not reported.</p> <p><u>Mean years in education</u>: not reported.</p> <p><u>Ethnicity</u>: not reported.</p>
Index tests	<p><u>Index test</u>: RUDAS</p> <p><u>Other tests</u>: none</p> <p>Tests were conducted by video call and face-to-face by assessors (advanced trainees in geriatric medicine) trained by a neuropsychologist. There was a calibration period to ensure consistency in the scoring. It was not reported whether assessors were blinded to the knowledge of the reference standard. Participants</p>

Wong 2012 (Continued)

were allocated at random to receive either the video-call or face-to-face assessment first. The RUDAS was modified for video call in the following ways: the video call provided a mirror image and a reference point for "left" was used to ensure accurate scoring, hand actions were performed at chest rather than desk height, and a felt pen rather than pencil was used.

A test threshold of 23 was prespecified for a diagnosis of dementia.

Target condition and reference standard(s)	<p><u>Target condition</u>: dementia (moderate or severe cognitive impairment).</p> <p><u>Reference standard</u>: the treating clinician (advanced trainees in geriatric medicine) diagnosed dementia. The diagnostic criteria or procedure for assessing the reference standard used were not reported. The reference standard was conducted independently of the index test.</p>
Flow and timing	<p>There was a minimum of 2 hours between video-call and face-to-face assessments, but not > 24 hours between assessment.</p> <p>No dropouts were reported.</p> <p>Data on true and false positives and negatives reported in the paper, and sensitivity and specificity were calculated in Review Manager 5 (true positive: 8, false positive: 3, true negative: 29, false negative: 2) (Review Manager 2020).</p>
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Wong 2012 (Continued)

Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition? Unclear

Were the reference standard results interpreted without knowledge of the results of the index tests? Yes

Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk

Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern

DOMAIN 4: Flow and Timing

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

Did all patients receive the same reference standard? Yes

Were all patients included in the analysis? Yes

Could the patient flow have introduced bias? Low risk

Zhou 2004
Study characteristics

Patient Sampling	Participants recruited from longitudinal ageing study. Study included in-patients and outpatients at the memory clinic at Xuanwu hospital in Beijing, China. <u>Sampling procedure:</u> not stated. <u>Inclusion criteria:</u> patients of the memory clinic at Xuanwu hospital in Beijing. <u>Exclusion criteria:</u> severe sensorimotor deficit, e.g. severe hearing deficit, delirium, or other acute physical conditions. Other diagnosed psychiatric diseases, e.g. depression.
Patient characteristics and setting	<u>Total number:</u> 132 <u>Mean age (years):</u> dementia 76 (SD 6.2), normal cognition: 74 (SD 5.5) <u>Sex (% female):</u> 56% (74) overall, dementia 56.9% (37), normal cognition: 55% (37)

Zhou 2004 (Continued)

Mean years in education: dementia 7.6 (SD 5.2), normal cognition: 7.3 (SD 4.5)

Ethnicity: not stated.

Index tests

Index test: IQCODE

Other tests: BRDS, IMCT

Tests administered by telephone by trained personnel.

IQCODE and BRDS were undertaken first and then IMCT completed with an informant present to ensure aids such as calendar or notes were not used. Informants were also used to verify participant's responses. 2 participants underwent 2 assessments for inter-rater reliability and 20 underwent subsequent testing for test-retest reliability.

Target condition and reference standard(s)

Target condition: dementia

Reference standard: patient history, examination, neuropsychological testing and laboratory tests done in face-to-face assessments. Diagnoses made by senior neurologists according to DSM-IV criteria.

Flow and timing

All participants received index tests and reference standards. Order of administration was randomised. Interval between face-to-face and telephone assessments was 10–19 days.

Comparative

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Zhou 2004 (Continued)

Could the conduct or interpretation of the index test have introduced bias?

Low risk

Are there concerns that the index test, its conduct, or interpretation differ from the review question?

Low concern

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition? Yes

Were the reference standard results interpreted without knowledge of the results of the index tests? Yes

Could the reference standard, its conduct, or its interpretation have introduced bias?

Low risk

Are there concerns that the target condition as defined by the reference standard does not match the question?

Low concern

DOMAIN 4: Flow and Timing

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

Did all patients receive the same reference standard? Yes

Were all patients included in the analysis? Yes

Could the patient flow have introduced bias?

Low risk

ACE: Addenbrooke's Cognitive Examination; AD: Alzheimer's disease; ALFI-MMSE: Adult Lifestyles and Function Interview Mini-Mental State Examination; AQ: Alzheimer's Questionnaire; BNPS: Brief Neuropsychiatric Screening test; BRDS: Blessed-Roth Dementia Scale; CDR: Clinical Dementia Rating; DSM III-R: Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; IMCT: Information Memory Concentration Test; IQCODE: Informant Questionnaire on Cognitive Decline in the Elderly; MCI: mild cognitive impairment; MIS-t: Memory Impairment Screen-Telephone; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; NINCDS-ADRDA: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; RUDAS: Rowland Universal Dementia Assessment Scale; SD: standard deviation; SPMSQ: Short Portable Mental Status Questionnaire; TEXAS: Telephone version of the Executive Interview; TICS-m: Modified Telephone Interview for Cognitive Status.

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Abdolahi 2016	Long time interval between remote and in-person assessment.
Bentvelzen 2019	Long time interval between remote and in-person assessment.
Chumbler 1998	No full text available.
Ciemins 2009	Wrong population.
Cook 2009	Participants with mild cognitive impairment (CDR < 0.5) rather than dementia.

Study	Reason for exclusion
Crooks 2005	Informant-based assessment.
Cullum 2006	Same person administering both the in-person and remote assessment.
Fong 2009	Telephone assessment but conducted in-person.
Gatz 2002	Only those who screened positive were further evaluated and with a long time interval between assessments.
Go 1997	Informant-based assessment.
Graff-Radford 2006	Probable case-control study and participants with dementia and mild cognitive impairment, data not presented separately.
Kempen 2007	Same person administering both the in-person and remote assessment.
Knopman 2010	Same person administering both the in-person and remote assessment.
Larner 2021	Telephone assessment but conducted in-person.
Lipton 2003	Long time interval between remote and in-person assessment.
Manly 2011	Informant-based assessment combined with direct assessment and data not presented separately.
Martin-Khan 2012	Remote assessment of dementia rather than a cognitive test.
Myers 2016	Self-completed assessment.
Tremont 2016	Longitudinal study.
Vercambre 2010	Combined mild cognitive impairment and dementia cohort, data not provided separately.
Wadsworth 2018	Combined mild cognitive impairment and dementia cohort, data not provided separately.
Wong 2015	Remote assessment was significantly truncated and not considered to be comparable to the full in-person assessment.
Younes 2007	Self-administered test.
Zhou 2020	Remote assessment was significantly truncated and not considered to be comparable to the full in-person assessment.
Zietemann 2017	Participants with mild cognitive impairment (CDR < 0.5) rather than dementia.

CDR: Clinical Dementia Rating.

DATA

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

Table Tests. Data tables by test

Test	No. of studies	No. of participants
2 ALFI-MMSE	2	233
3 IMCT	1	132
4 SPSMQ	1	100
5 Tele-Free-Cog	1	108
7 RUDAS	1	42

Test 2. ALFI-MMSE

ALFI-MMSE

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Cammozzato 2011	66	17	0	50	1.00 [0.95, 1.00]	0.75 [0.63, 0.84]		
Roccaforte 1992	44	0	22	34	0.67 [0.54, 0.78]	1.00 [0.90, 1.00]		

Test 3. IMCT

IMCT

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Zhou 2004	52	13	13	54	0.80 [0.68, 0.89]	0.81 [0.69, 0.89]		

Test 4. SPSMQ

SPSMQ

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Roccaforte 1994	27	1	39	33	0.41 [0.29, 0.54]	0.97 [0.85, 1.00]		

Test 5. Tele-Free-Cog

Tele-Free-Cog

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Burns 2020	45	0	7	56	0.87 [0.74, 0.94]	1.00 [0.94, 1.00]		

Test 7. RUDAS

RUDAS

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Wong 2012	8	3	2	29	0.80 [0.44, 0.97]	0.91 [0.75, 0.98]		

ADDITIONAL TABLES

Table 1. Characteristics of studies comparing remote with in-person assessment

Study	Patient selection	Applicability concerns	Index test	Reference standard	Flow and timing
Telephone assessments					
Arnold 2009	<p>Older adults (> 65 years) recruited from the Cardiovascular Health Study.</p> <p><u>Sampling procedure:</u> random sampling.</p> <p>Participants randomly selected to receive 3MSE, TICS and IQCODE stratified by education, and prior 3MSE scores. Additional participants outside of strata were recruited to reach the target sample sizes.</p> <p><u>Inclusion criteria:</u> aged ≥ 65 years, not wheelchair bound, able to provide informed consent, did not require a proxy respondent, were not under active treatment for cancer and planned to remain in the area for ≥ 3 years.</p> <p><u>Exclusion criteria:</u> not reported.</p>	<p>Participants were community dwelling older adults (aged > 65 years) enrolled in a USA cohort.</p> <p><u>Total number of participants:</u> 405, 290 with all 3 cognitive measures, 74 with TICS-m and 3MSE only, and 41 with IQCODE and 3MSE only.</p> <p><u>Mean age (years):</u> 80.5 (range 70–102).</p> <p><u>Sex (% female):</u> 62%.</p> <p><u>Education:</u> 52.2% > 10 years' education.</p> <p><u>Ethnicity:</u> 28% African American.</p>	<p><u>Index test:</u> TICS-m.</p> <p><u>Other tests conducted:</u> IQCODE.</p> <p>Used the 40-point version of the TICS. No information provided on the assessors.</p> <p>3MSE conducted annually at in-person clinic visits.</p> <p>TICS-m and IQCODE evaluated against the 3MSE within 30 days.</p> <p>No information provided on the administration of the tests.</p> <p>TICS-m and IQCODE attempted for participants who did not have an in-person clinic visit. Thresholds prespecified for 3MSE (< 80), and TICS (< 25).</p> <p>Unclear whether the 3MSE and TICS were conducted independently of one another.</p> <p>Assessments conducted in English.</p>	<p><u>Reference standard:</u></p> <p>Discriminative ability of the TICS compared to 3MSE score.</p> <p>No information provided on the assessors or conduct of the 3MSE.</p>	<p>The TICS-m was conducted within 30 days of the 3MSE.</p> <p>405 participants recruited, dropouts not reported.</p> <p>364 participants had both the 3MSE and TICS-m.</p>
Baccaro 2015	<p>Participants recruited from the Study of Stroke Mortality and Morbidity in Adult (EMMA study).</p>	<p><u>Total number of participants:</u> 61</p> <p><u>Mean age (years):</u> 61.9 (range 22–87)</p>	<p><u>Index test:</u> TICS-m (modified Brazilian-Portuguese version).</p>	<p><u>Reference standard:</u> MMSE.</p>	<p>TICS-m conducted at 1 and 2 weeks after the in-person evaluation.</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

<p>Conducted in Brazil.</p> <p><u>Sampling procedure:</u> not reported.</p> <p><u>Inclusion criteria:</u> clinical diagnosis of stroke; age ≥ 35 years; Portuguese speaker.</p> <p><u>Exclusion criteria:</u> moderate-to-severe neurological disease; alcohol and substance abuse; aphasia and other significant acute medical conditions.</p>	<p><u>Sex (% female):</u> 36.1%</p> <p><u>Education:</u> 14.8% illiterate, 42.6% < 7 years, 19.7% 8–11 years, 23% > 11 years</p> <p><u>Ethnicity:</u> 44.3% white, 4.9% black, 39.3% mixed, 11.5% Asiatic</p>	<p>Thresholds prespecified as: ≤ 21 (no formal education), ≤ 22 (low education), ≤ 23 (intermediate education), ≤ 24 (high education).</p> <p>However, in the reported results an optimal threshold was used at < 14.</p> <p>Test translated and culturally adapted in a subsample of 30 non-clinical participants who were accompanying patients at the emergency department, or from the community.</p> <p><u>Other tests:</u> Hamilton Depression Rating Scale.</p>	<p>All tests carried out by 2 psychologists but unclear whether the index test and reference standard were independent of one another.</p>	<p>61 participants included in the analysis (of whom 14 had cognitive impairment).</p> <p>504 participants enrolled in the EMMA study, 122 (24.2%) met the exclusion criteria; 165 (32.7%) were not found or not reached on time to participate, 85 (16.9%) died before 6 months of follow-up, and 42 (8.4%) eligible patients refused to participate. 90 (17.9%) patients were invited to the initial interview, 17 (0.3%) did not attend the interview, and 12 people (0.2%) were lost along the follow-up.</p>	
<p>Castanho 2015</p>	<p>Participants recruited while in the waiting area of a health centre for their annual check-up.</p> <p><u>Sampling procedure:</u> convenience sample, stratified by age and sex.</p> <p><u>Inclusion criteria:</u> not reported.</p> <p><u>Exclusion criteria:</u> inability to consent, inability to read or write, dementia, diagnosed neuropsychiatric, neurodegenerative disorder.</p>	<p>Study undertaken in Portugal.</p> <p>Participants were equally distributed between urban and rural areas.</p> <p><u>Total sample:</u> 142 participants</p> <p><u>Mean age (years):</u> 67.5 (range 52–84)</p> <p><u>Sex (% female):</u> 56.3%</p> <p><u>Education:</u> median 4 (range 0–17) years, < 4 years 69.7%, ≥ 4 years 30.3%</p>	<p><u>Index test:</u> TICS-m (Portuguese version).</p> <p><u>Other tests:</u> none.</p> <p>Index tests were conducted without the knowledge of the reference standard.</p> <p>TICS-m was translated into Portuguese by 2 psychologists and a biochemist, and back translated by a further psychologist and linguist. The translated version was further reviewed in multidisciplinary team discussions, in a pilot survey, and a pilot study of 33 com-</p>	<p><u>Reference standard:</u> MMSE, MoCA.</p> <p>Reference standard was conducted independently of the index tests by a trained psychologist.</p>	<p>Index tests conducted within 1 month of in-person assessment.</p> <p>All participants completed the TICS-m and MMSE.</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

		<u>Ethnicity</u> : not reported.	community-dwelling older adults.		
			The final version of the TICS-m-PT comprised 13 items with a maximum score of 39 points.		
			A psychologist conducted the TICS-m.		
			Test thresholds were not prespecified and optimal cut-offs were calculated using the study data.		
Christodoulou 2016	<p>Participants recruited from outpatient clinics.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: diagnosed with sporadic ALS, progressive muscular atrophy, suspected primary lateral sclerosis, bulbar palsy, or predominantly upper motor neuron disease; with a disease duration of ≤ 18 months after symptom onset; aged ≥ 20 years; reliable family carer who could give independent informed consent for providing information and assisting with testing; fluent in English; capacity to consent.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p>Study undertaken in the USA.</p> <p><u>Total number of participants</u>: 31</p> <p><u>Mean age (years)</u>: 62 (SD 8)</p> <p><u>Sex (% female)</u>: 35%</p> <p><u>Education</u>: 68% college degree, 6% some college, 3% associates degree, 3% high school</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: University of California Screening Battery (UCSF).</p> <p><u>Other tests</u>: TICS.</p> <p>Cognitive assessment was modified to be delivered by telephone by removal of the visual assessments, saccades and anti-saccades were replaced with tapping commands.</p> <p>All other components were the same as the original assessment. Assessments were conducted in English.</p> <p>Participants were assigned to 2 groups to either undergo the in-person or telephone assessment first.</p> <p>Index test and reference standard conducted by the same trained interviewer.</p> <p>Test thresholds were not prespecified.</p>	<p><u>Reference standard</u>: In-person assessment of UCSF and MMSE.</p>	<p>Index test conducted within 2 weeks of the reference standard.</p> <p>Dropouts not reported.</p>
Crooks 2007	<p>908 adults aged > 65 years randomly selected from Kaiser Permanente Southern California and recruited by letter invitation.</p> <p><u>Sampling procedure</u>: random, equal numbers of males, females and ethnic groups were targeted.</p>	<p>Community dwelling adults in the USA.</p> <p><u>Total number</u>: 211</p> <p><u>Age (years)</u>: 32% 60–69, 28% 70–74, 25% 75–79, 10% 80–84, 5% ≥ 85</p>	<p><u>Index test</u>: CALLS.</p> <p><u>Other tests</u>: none.</p> <p>Unclear whether index test was conducted without knowledge of the reference standard.</p>	<p><u>Reference standard</u>: MMSE.</p> <p>Lay interviewers with degree were trained by neuropsychologists and su-</p>	<p>Mean time between tests 16.27 (range 1–60) days.</p> <p>152 participants excluded from study, and not all partici-</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

	<p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: language barrier; aged < 65 years; illness; severe hearing problem and relocation out of area.</p>	<p><u>Sex (% female)</u>: 51%</p> <p><u>Education</u>: 7% high school, 20% high school graduate, 39% some college, 34% college or higher</p> <p><u>Ethnicity</u>: 21% Asian; 23% African American; 19% Hispanic, 36% white</p>	<p>No information was provided on the assessors of the index test.</p> <p>Test threshold was not prespecified.</p>	<p>pervised by project staff to perform the test battery.</p> <p>197 participants received MMSE.</p>	<p>pants received MMSE.</p>
<p>Dellasega 2001</p>	<p>Participants recruited from a primary care office.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: if unable to complete both morning and afternoon sessions.</p>	<p>Participants scheduled for geriatric assessment or follow-up visit.</p> <p><u>Total number</u>: 12</p> <p><u>Mean age (years)</u>: 81.3 (SD 1)</p> <p><u>Sex (% female)</u>: 58%</p> <p><u>Education</u>: 25% less than high school, 8% some high school, 42% high school, 25% college</p> <p><u>Ethnicity</u>: 100% Caucasian</p>	<p><u>Index test</u>: OMC telephone.</p> <p>Index test administered by the research assistant (medical student or geriatric nurse practitioner) by telephone following the in-person assessment.</p> <p>Thresholds not pre-specified.</p>	<p><u>Reference standard</u>: OMC in-person.</p> <p>Administered by research assistant (medical student or geriatric nurse practitioner) in-person.</p> <p>In addition to administering the OMC, the research assistants made independent clinical judgments about the participants' mental status. This impression was a subjective rating based on interactions with the participants.</p>	<p>Reference standard and index test conducted on the same day.</p> <p>Dropouts not reported.</p>
<p>Garre-Olmo 2008</p>	<p>Outpatients with a diagnosis of dementia were recruited.</p> <p><u>Sampling procedure</u>: consecutive.</p> <p><u>Inclusion criteria</u>: informed consent, telephone in the place of residence.</p> <p><u>Exclusion criteria</u>: hearing deficit, serious illness, MMSE score < 5</p>	<p><u>Total number</u>: 141</p> <p><u>Mean age (years)</u>: 76.6 (SD 6.9)</p> <p><u>Sex (% female)</u>: 69%</p> <p><u>Education</u>: 14% illiterate, 36% 1–6 years, 49% > 7 years</p>	<p><u>Index test</u>: telephone MMSE.</p> <p>Order of the remote and in-person assessments were randomised.</p>	<p><u>Reference standard</u>: in-person MMSE.</p> <p>No information on conduct of the in-person MMSE.</p>	<p>Mean interval of 10 days between telephone and in-person assessments.</p> <p>There were no dropouts, 29 patients did not agree or meet</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

		<u>Ethnicity</u> : not reported.			the inclusion criteria.
Kennedy 2014	<p>Participants in a cohort study (University of Alabama at Birmingham Study of Aging II).</p> <p><u>Sampling procedure</u>: random, stratified.</p> <p><u>Inclusion criteria</u>: aged ≥ 75 years, living independently, able to schedule study appointments and answer questions independently, able to complete index and reference tests.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p>Community dwelling older adults.</p> <p><u>Total number</u>: 402</p> <p><u>Mean age (years)</u>: 81.6 (SD 4.8)</p> <p><u>Sex (% female)</u>: 58%</p> <p><u>Mean education</u>: 8% elementary school, 50% high school, 3% some college, 10% college graduate</p> <p><u>Ethnicity</u>: 65% Caucasian</p>	<p><u>Index test</u>: MMSET.</p> <p>No information provided on the assessor of the index test.</p> <p>Unclear whether they were conducted independently of the reference standard.</p> <p>Test threshold was not prespecified.</p>	<p><u>Reference standard</u>: MMSE, SIS.</p> <p>Trained interviewers conducted the assessments.</p> <p>Unclear whether the reference standard was conducted without the knowledge of the index test results.</p>	<p>Time interval between index test and reference standard not specified.</p> <p>419 participants underwent baseline assessments, 17 did not have a telephone assessment at follow-up.</p> <p>402 were included in the final sample.</p>
Lanska 1993	<p>No information provided on the recruitment procedure.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: Hachinski score < 4, EEG and brain imaging consistent with AD.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p>All participants had a diagnosis of probable AD.</p> <p><u>Total number</u>: 30</p> <p><u>Median age (years)</u>: 76 (range 59–88)</p> <p><u>Sex (% female)</u>: 73%</p> <p><u>Median education</u>: 12 (range 4–20) years</p> <p><u>Ethnicity</u>: not reported</p>	<p><u>Index test</u>: TAMS.</p> <p>The index test was administered by a trained psychometrist independently of the reference standard.</p>	<p><u>Reference standard</u>: MMSE and ADAS.</p> <p>Scales administered by a trained psychometrist, who was blinded to the index test results.</p>	<p>Reference standard administered within 1 week of index test.</p> <p>2 participants were unable to be assessed by telephone and were assigned a score of 0.</p>
Matrisch 2012	<p>Participants recruited from 9 primary care practices.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: aged ≥ 70 years, absence of acute illness, ability to visit the general practitioner's practice, availability by telephone and written consent.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p><u>Total number</u>: 197</p> <p><u>Median age (years)</u>: 78.5 (SD 4.1)</p> <p><u>Sex (% female)</u>: 58%</p> <p><u>Education</u>: not reported.</p>	<p><u>Index test</u>: TICS-m.</p> <p>TICS-m was carried out independently of the reference standard.</p> <p>No information provided on the assessor.</p>	<p><u>Reference standard</u>: MMSE.</p> <p>MMSE administered in the primary care practice.</p> <p>Unclear if this was conducted independently of the reference standard.</p>	<p>Interval between the index test and reference standard was not reported.</p> <p>Dropouts were not reported.</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

Metitieri 2001	<p>Admissions to an Alzheimer's unit.</p> <p><u>Sampling procedure</u>: consecutive.</p> <p><u>Inclusion criteria</u>: admission to the Alzheimer's Unit.</p> <p><u>Exclusion criteria</u>: deafness, severe aphasia, acute conditions and advanced dementia (classified as CDR of 4 or 5).</p>	<p><u>Total number</u>: 104</p> <p><u>Mean age (years)</u>: 77.2 (SD 8.1)</p> <p><u>Sex (% female)</u>: 76%</p> <p><u>Mean education</u>: 5.2 (SD 2.3) years</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: Itel-MMSE.</p> <p>Itel-MMSE was adapted from the ALFI-MMSE.</p> <p>No information on the adaptation process. Itel-MMSE is an Italian language, telephone-adapted version of the MMSE.</p> <p>Participants received 3 different administrations of the Itel-MMSE – 2 by the same assessor within 1 week for test-retest reliability and a third by a different assessor to measure inter-rater reliability.</p>	<p><u>Reference standard</u>: MMSE face-to-face.</p> <p>Unclear whether different assessors completed the index test and the reference standard.</p>	<p>Time interval between administration of index test and reference standard unclear.</p> <p>2 administrations of the index test were within 1 week of each other.</p>
Monteiro 1998	<p>Participants in a cohort study.</p> <p><u>Sampling method</u>: not reported.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: participants involved in initial evaluation study.</p>	<p><u>Total number</u>: 30</p> <p><u>Mean age (years)</u>: 77.6 (SD 7.8)</p> <p><u>Sex (% female)</u>: 57%</p> <p><u>Mean education (years)</u>: not reported.</p>	<p><u>Index tests</u>: telephone assessments of MMSE, BCRS.</p> <p><u>Other tests</u>: Global Deterioration Scale, Functional Assessment Staging, Behavioural Pathology in Alzheimer's Disease Rating Scale.</p> <p>Index test conducted independent of the reference standard by a clinician.</p>	<p><u>Reference standard</u>: in-person MMSE and BCRS.</p> <p>Reference standard was conducted by a second clinician who was blinded to the telephone assessment.</p>	<p>Mean time between index and reference tests was 7.2 (range 2–30) days.</p> <p>2 participants did not undergo telephone interview.</p>
Newkirk 2004	<p>Participants with a diagnosis of AD in a cohort study.</p> <p><u>Sampling method</u>: not reported.</p> <p><u>Inclusion criteria</u>: aged 55–90 years with a diagnosis of AD, and MMSE score > 5, and a carer willing to participate in the longitudinal study.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p>Participants with a diagnosis of AD.</p> <p><u>Total number</u>: 46</p> <p><u>Mean age (years)</u>: 76.5</p> <p><u>Sex (% female)</u>: 52%</p> <p><u>Mean education (years)</u>: ≤ 8 years: 2%, 9–12 years: 26%, > 12 years: 72%</p> <p><u>Ethnicity</u>: 87% Caucasian</p>	<p><u>Index tests</u>: ALFI-MMSE.</p> <p>Index test administered by a clinical research assistant.</p> <p>Not reported whether this was independent of the in-person assessment.</p>	<p><u>Reference standard</u>: in-person MMSE.</p> <p>MMSE delivered in-person by a clinician.</p> <p>Not reported if this was independent of the telephone assessment.</p>	<p>Interval between the in-person and telephone assessment was 1–35 days.</p> <p>7 people missing from the analysis.</p>
Plassman 1994	<p>Participants in a cohort study.</p>	<p>Male twin veterans referred for suspected AD.</p>	<p><u>Index tests</u>: TICS-m.</p> <p>TICS-m administered by trained, lay inter-</p>	<p><u>Reference standard</u>: in-person MMSE.</p>	<p>Mean 18 days between telephone and in-</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

	<p><u>Sampling method</u>: not reported.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p><u>Total number</u>: 67</p> <p><u>Mean age (years)</u>: 69 (SD 3)</p> <p><u>Sex (% female)</u>: 0%</p> <p><u>Mean education (years)</u>: 13 (4)</p> <p><u>Ethnicity</u>: not reported.</p>	<p>viewers, blinded to the knowledge of the reference standard.</p>	<p>In-person assessment occurred before the telephone and therefore was blinded to the results of this.</p>	<p>person assessment.</p> <p>14 participants missing from the analyses.</p>
Video-call assessments					
Ball 1993	<p>Participants aged ≥ 65 years recruited from inpatients at an old age psychiatry unit.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p>Participants were aged > 65 years of age from an inner-city UK based psychiatry service.</p> <p><u>Total number</u>: 11</p> <p><u>Age (years)</u>: range 22–58</p> <p><u>Sex (% female)</u>: 18%</p> <p><u>Education</u>: not reported.</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: CAMCOG.</p> <p><u>Other tests</u>: none.</p> <p>Index test conducted independently of reference standard by psychologists. Order of remote and in-person assessments was randomised.</p>	<p><u>Reference standard</u>: in-person assessment of the CAMCOG.</p> <p>Reference standard was conducted independently of the index tests by psychologists.</p>	<p>48-hour interval between remote and in-person assessments.</p> <p>1 dropout from the study.</p> <p>Test accuracy data not reported.</p>
Carotenuto 2018	<p>Participants with diagnosis of AD recruited from the Neurodegenerative Diseases Unit in Naples, Italy.</p> <p><u>Sampling procedure</u>: random.</p> <p><u>Inclusion criteria</u>: aged > 50 years, MMSE score 12–24, > 5 years' education, good hearing and vision, living or in contact with a carer.</p> <p><u>Exclusion criteria</u>: decompensated heart disease, chronic renal failure, severe liver failure, uncorrected thyroid disorder, cancer, major depression, different diagnosis of AD.</p>	<p>Outpatients from a specialist clinic in Italy aged > 50 years.</p> <p><u>Total number</u>: 28</p> <p><u>Age (years)</u>: 73.8 (SD 7.5)</p> <p><u>Sex (% female)</u>: 71%</p> <p><u>Education (years)</u>: 7.6 (SD 4.1)</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: MMSE and ADAS-Cog.</p> <p><u>Other tests</u>: none.</p> <p>Remote tests administered independently of the in-person assessment by a trained and experienced psychologist.</p>	<p><u>Reference standard</u>: in-person assessment of the MMSE and ADAS-Cog.</p> <p>Tests administered independently of the remote assessments by a trained and experienced psychologist.</p>	<p>Interval between remote and in-person assessments was 2 weeks.</p> <p>No missing data.</p> <p>Neither MMSE nor ADAS-Cog scores differed significantly between remote and in-person testing.</p>
Cullum 2014	<p>Participants recruited from the Alzheimer's Disease Center at the University of Texas Southwestern Medical Center</p>	<p>Participants were both urban and rural dwelling.</p>	<p><u>Index test</u>: MMSE.</p> <p><u>Other tests</u>: neuropsychological battery.</p>	<p><u>Reference standard</u>: in-person assessment of the index tests.</p>	<p>No missing data and remote and in-person assessments were conduct-</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

	in Dallas, and a satellite clinic in Talihina, Oklahoma. <u>Sampling procedure</u> : not reported. <u>Inclusion criteria</u> : fluent in English (primary language), adequate hearing and vision. <u>Exclusion criteria</u> : not reported.	<u>Total number</u> : 202 <u>Age (years)</u> : 68.5 (SD 9.5) <u>Sex (% female)</u> : 63% <u>Education (years)</u> : 14.1 (SD 2.7) <u>Ethnicity</u> : not reported.	Participants randomised to either in-person or video call first. Tests performed by trained research assistants or psychometrists. No information provided on blinding of the index tests.	No information reported on the blinding of the reference standard assessment. For rural participants, some of the reference standard was administered remotely, in addition to in-person assessment.	ed on the same day.
Hwang 2022	Participants recruited from 2 Australian aged care assessment services, and from a multicultural community service. <u>Sampling procedure</u> : not reported. <u>Inclusion criteria</u> : community dwelling adults aged > 65 years from a culturally or linguistically diverse background, requiring an interpreter, with carer or individual reported cognitive impairment. <u>Exclusion criteria</u> : not reported.	Participants were recruited from culturally and linguistically diverse backgrounds with limited English language proficiency. <u>Total number</u> : 90 <u>Age (years)</u> : 76 <u>Sex (% female)</u> : 64% <u>Education</u> : 47% at ≤ 8 year <u>Ethnicity</u> : not reported.	<u>Index test</u> : RUDAS. <u>Other tests</u> : GDS. Index test was delivered in-person but with an interpreter via video call. No information provided on the blinding of the index tests.	<u>Reference standard</u> : in-person assessment of the index test. Reference standard was conducted by an allied health professional with an interpreter. No information provided on the blinding of the reference standard.	7–14 days between remote and in-person assessments. 19 (17%) participants did not complete both assessments. RUDAS scores were similar between video call (27) and in-person assessment (28), with no significant difference on paired t-test, and a correlation coefficient of 0.73 between the 2 modalities.
Loh 2004	Participants were inpatients on a postoperative rehabilitation unit for fractured neck of femur or acute medical unit presenting to geriatricians with a range of conditions (dementia, delirium, depression). <u>Sampling procedure</u> : not reported. <u>Inclusion criteria</u> : not reported. <u>Exclusion criteria</u> : not reported.	Inpatients aged ≥ 65 years. <u>Total number</u> : 20 <u>Age (years)</u> : 82 (range: 72–95) <u>Sex (% female)</u> : 80% <u>Education</u> : not reported. <u>Ethnicity</u> : not reported.	<u>Index test</u> : SMMSE. <u>Other tests</u> : GDS. Assessments were conducted by an advanced trainee in geriatric medicine. The order of in-person and remote assessments was randomised. Remote tests were conducted independently of the in-person assessment.	<u>Reference standard</u> : SM-MSE. Assessments conducted by an advanced trainee in geriatric medicine. In-person assessment conducted independently of the remote tests.	Interval between the in-person and remote assessment was not reported. No dropouts reported.

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

Loh 2007	<p>Rural dwelling people in Australia, referred by general practitioners for symptoms of cognitive impairment.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: fluent in English with no significant visual or sensory impairment that would prevent participation in a video call.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p>Rural dwelling participants in Australia referred for assessment of cognitive impairment.</p> <p><u>Total number</u>: 20</p> <p><u>Mean age (years)</u>: 79 (range 67–89)</p> <p><u>Sex (% female)</u>: 55%</p> <p><u>Education</u>: not reported.</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: MMSE.</p> <p><u>Other tests</u>: GDS, IQ-CODE, activities of daily living, instrumental activities of daily living.</p> <p>Index test was delivered via video call by geriatricians. Assessors were blinded to reference standard.</p> <p>Test threshold not pre-specified.</p> <p>Allocation to the order of the index test and reference standard was alternate.</p>	<p><u>Reference standard</u>: MMSE.</p> <p>Reference standard administered by a specialist geriatrician (single clinician assessment). Assessors were blinded to the knowledge of the index test.</p>	<p>Remote assessments were conducted within 1 week of the in-person assessment.</p> <p>No dropouts reported.</p>
Menon 2001	<p>Participants aged ≥ 60 years and inpatients at an acute medical unit or a geriatric evaluation and management unit.</p> <p><u>Sampling procedure</u>: convenience sample.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: too unwell to participate.</p>	<p><u>Total number</u>: 12</p> <p><u>Age</u>: not reported.</p> <p><u>Sex</u>: not reported.</p> <p><u>Education</u>: not reported.</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: SPMSE.</p> <p><u>Other tests</u>: GDS, Hamilton Depression Scale.</p> <p>Video-call assessments conducted by geropsychiatry fellows independently of the in-person assessment.</p>	<p><u>Reference standard</u>: in-person SP-MSE.</p> <p>In-person assessment conducted by a geropsychiatry fellow independently of the remote assessment.</p>	<p>Remote and in-person assessments conducted on the same day, or within 1 week.</p> <p>No dropouts reported.</p>
Vahia 2015	<p>Participants were monolingual or bilingual Spanish-speaking who were referred for assessment for suspected cognitive impairment.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: severe concurrent medical illness, major psychiatric disorder, sensory impairments or previous neurological impairments.</p>	<p>Participants were aged > 65 years with suspected cognitive impairment.</p> <p><u>Total number</u>: 17</p> <p><u>Age (years)</u>: 70–71</p> <p><u>Sex (% female)</u>: 18–27%</p> <p><u>Education</u>: 5–5.9 years</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: MMSE.</p> <p><u>Other tests</u>: neuropsychological battery.</p> <p>Tests undertaken by 2 trained clinical evaluators, independently of the in-person assessment.</p> <p>Order of remote and in-person assessments was randomised.</p>	<p><u>Reference standard</u>: in-person assessment of the MMSE.</p> <p>Tests undertaken by 2 trained clinical evaluators, independently of the remote assessment.</p>	<p>10 participants did not complete the study.</p> <p>Interval of 2 weeks.</p>
Wadsworth 2016	<p>Participants were recruited from the Alzheimer's Disease Center at the University of Texas.</p> <p><u>Sampling procedure</u>: not reported.</p>	<p>Cognitively impaired participants included both MCI and AD.</p> <p><u>Total sample</u>: 197</p>	<p><u>Index test</u>: MMSE.</p> <p><u>Other tests</u>: neuropsychological battery.</p> <p>Tests completed by experienced examiners.</p>	<p><u>Reference standard</u>: in-person assessment of the index test.</p>	<p>Most tests were completed within 2.5 hours of each other, 2 were completed at 7 and 14 days.</p>

Table 1. Characteristics of studies comparing remote with in-person assessment (Continued)

	<p><u>Inclusion criteria</u>: fluent in English with no significant visual or sensory impairment that would prevent participation in a video call.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p>(78 cognitively impaired)</p> <p><u>Mean age (years)</u>: healthy: 66.1 (SD 9.2), dementia 72.7 (SD 8.4)</p> <p><u>Sex (% female)</u>: healthy: 75%, dementia: 46%</p> <p><u>Education (years)</u>: healthy: 13.9 (SD 2.5), dementia: 14.6 (SD 3.1)</p> <p><u>Caucasian</u>: healthy: 50%, dementia: 62%</p>	<p>Test order was counter-balanced across participants. No information was provided on blinding.</p>	<p>No information provided on blinding.</p>	<p>Missing data occurred in up to 20 participants (10%).</p> <p>Test scores differed significantly between remote and in-person assessment only for the category fluency test.</p>
Wong 2011	<p>Participants recruited from an inpatient Geriatric rehabilitation unit in a large teaching hospital.</p> <p><u>Sampling procedure</u>: not reported.</p> <p><u>Inclusion criteria</u>: not reported.</p> <p><u>Exclusion criteria</u>: not reported.</p>	<p><u>Total sample</u>: 42</p> <p><u>Mean age (years)</u>: 74.8</p> <p><u>Sex</u>: not reported.</p> <p><u>Education</u>: not reported.</p> <p><u>Ethnicity</u>: not reported.</p>	<p><u>Index test</u>: RUDAS.</p> <p><u>Other tests</u>: none.</p> <p>Assessment order and allocation to 1 of 2 assessing doctors was randomised. Threshold of 23 used to determine a diagnosis of dementia.</p>	<p><u>Reference standard</u>: in-person assessment of the index test.</p>	<p>Time interval between the remote and in-person assessments not reported.</p> <p>In-person and video-call assessments correlated with a coefficient of 0.79, with no significant differences. Agreement between conditions was 71%, ± 2 points in 30 cases.</p> <p>Agreement for a diagnosis of dementia was 88%.</p>

3MSE: modified Mini-Mental State Examination; AD: Alzheimer's disease; ADAS: Alzheimer's Disease Assessment Scale; ADAS-Cog: Alzheimer's Disease Assessment Cognitive Subscale; ALS: amyotrophic lateral sclerosis; BCRS: Brief Cognitive Rating Scale; CALLS: Cognitive Assessment for Later Life Status; CAMCOG: Cambridge Cognition Examination; EEG: electroencephalogram; GDS: Geriatric Depression Scale; IQCODE: Informant Questionnaire on Cognitive Decline in the Elderly; MMSE: Mini-Mental State Examination; MMSET: Mini-Mental State Examination Telephone; MoCA: Montreal Cognitive Assessment; OMC: Ottawa Mental State Exam; RUDAS: Rowland Universal Dementia Assessment Scale; SD: standard deviation; SIS: six-Item Screener; TAMS: Telephone-Assessed Mental State; TICS: Telephone Interview for Cognitive Status; TICS-m: Modified Telephone Interview for Cognitive Status; TICSm-PT: Telephone Interview for Cognitive Status-Modified – Portuguese version; UCSF: University of California San Francisco.

Table 2. Main findings of studies comparing remote with in-person equivalent tests

Study	Comparison	Number included	ICC or kappa (k)	Difference in scores	Other
-------	------------	-----------------	------------------	----------------------	-------

Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia (Review)

51

Table 2. Main findings of studies comparing remote with in-person equivalent tests (Continued)
 (95% CI)

Telephone call assessments					
Arnold 2009	TICS-m vs F2F 3MSE	364	NR	NR	Sensitivity: 80%, specificity: 76% against 3MSE (threshold < 80)
Baccaro 2015	TICS-m vs F2F MMSE	61	NR	NR	Sensitivity: 91.5%, specificity: 71.4% against MMSE (education adjusted thresholds of 21–24)
Castanho 2015	TICS-m vs F2F MMSE	142	NR	NR	Sensitivity: 90.6%, specificity: 73.7% (threshold < 13.5)
Christodoulou 2016	UCSF vs F2F others	31	ALS-CBS: 0.50 (1.00 to 1.11) COWAT: 0.34 (0.90 to 0.99) WVFI: 0.76 (0.95 to 1.32)	NR	NR
Crooks 2007	CALLS vs F2F MMSE	211	NR	NR	NR
Dellasega 2001	OMC vs F2F OMC	12	NR	NR	Agreement between OMC scores in-person and by telephone was 100%
Garre-Olmo 2008	MMSE vs F2F MMSE	141	Full: 0.87 (0.83 to 0.91) Subscale: 0.86 (0.80 to 0.90)	MD 3.5 MMSE F2F 18.9 (SD 4.9) MMSET 15.4 (4.7) 22-item MMSE F2F 13 (5–22) MMSET 13 (3–22)	Remote scores lower
Kennedy 2014	MMSET vs F2F MMSE	402	NR	NR	MMSET showed good internal consistency with MMSE (Cronbach α = 0.845) AUC = 0.73 (MMSET), AUC = 0.70 (MMSE)
Lanska 1993	TAMS vs F2F MMSE	30	NR	NR	NR
Matrisch 2012	TICS-m vs F2F MMSE	197	NR	NR	NR
Metitieri 2001	Itel-MMSE vs F2F MMSE	104	NR	NR	NR
Monteiro 1998	MMSE vs F2F MMSE	30	NR	NR	NR

Table 2. Main findings of studies comparing remote with in-person equivalent tests (Continued)

	BCRS vs F2F BCRS	30	NR	NR	NR
Newkirk 2004	ALFI-MMSE vs F2F MMSE	46	NR	MD 2.9 MMSE F2F 18.5 (SD 6.14) ALFI-MMSE 15.6 (SD 6.92) 22-item subscale: MMSE 12.10 (SD 4.97) ALFI-MMSE 13.28 (SD 6.00)	Remote scores lower (full) Remote scores higher (subscale)
Plassman 1994	TICS-m vs F2F MMSE	67	NR	NR	NR
Video-call assessments					
Ball 1993	MMSE VC vs F2F MMSE	11	NR	MD 0.4 MMSE F2F 23.8 MMSE VC 24.2	Remote scores higher
Carotenuto 2018	ADAS Cog VC vs F2F ADAS Cog	28	NR	MD 5.5 ADAS Cog F2F 28.6 (SD 19.3) ADAS Cog VC 34.1 (SD 17.4)	Remote scores higher
	MMSE VC vs F2F MMSE	28	NR	MD 0.8 MMSE F2F 19.6 (SD 3) MMSE VC 18.8 (SD 4.5)	Remote scores lower
Cullum 2014	MMSE VC vs F2F MMSE	202	0.91	MMSE VC 27.6 (SD 3.1) MMSE F2F 27.6 (SD 3.1)	Comparable scores
Hwang 2022	RUDAS vs F2F RUDAS	45	NR	MD -0.36 (-1.09 to 0.38) RUDAS F2F 28 (25-29) RUDAS VC 27 (25-28)	Remote scores lower
Loh 2004	MMSE VC vs F2F MMSE	20	NR	MD -0.3 (-3.9 to 4.5) MMSE F2F 24.3 (SD 4.9) MMSE VC 24 (SD 4.9)	Remote scores lower
Loh 2007	MMSE vs F2F MMSE	20	NR	MD 0.93 (-1.89 to 0.04)	Remote scores higher
Menon 2001	SPMSE VC vs F2F SPMSE	12	NR	Coefficient of variation F2F 63% VC 32%	NR

Table 2. Main findings of studies comparing remote with in-person equivalent tests (Continued)

Vahia 2015	MMSE VC vs F2F MMSE	17	NR	MMSE F2F Z score -1.02 MMSE VC Z score -0.73	Remote scores higher
Wadsworth 2016)	MMSE vs F2F MMSE	83	0.92	MMSE VC 27.5 (SD 2.7) MMSE F2F 27.7 (SD 2.4)	Comparable scores
Wong 2011	RUDAS VC vs F2F RUDAS	42	NR	MD 0.05 (SD 2.7)	NR

3MSE: modified Mini-Mental State Examination; ADAS-Cog: Alzheimer's Disease Assessment Cognitive Subscale; ALFI-MMSE: Adult Lifestyles and Function Interview Mini-Mental State Examination; ALS-CBS: Amyotrophic Lateral Sclerosis Cognitive Behavioural Screen; AUC: area under the curve; BCRS: Brief Cognitive Rating Scale; CALLS: Cognitive Assessment for Later Life Status; COWAT: Controlled Oral Word Association Test; F2F: face-to-face; ICC: intraclass correlation coefficient; Itel-MMSE: Italian version of the telephone MMSE; MD: mean difference; MMSE: Mini-Mental State Examination; MMSET: Mini-Mental State Examination Telephone; NR: not reported; OMC: Ottawa Mental State Exam; RUDAS: Rowland Universal Dementia Assessment Scale; SD: standard deviation; SPMSE: Short Portable Mental State Examination; TAMS: Short Portable Mental State Examination; TICS-m: Modified Telephone Interview for Cognitive Status; UCSF: University of California Screening Battery; VC: videoconferencing; WVFI: Written Verbal Fluency Index.

APPENDICES

Appendix 1. Sources searched and search strategies

Source	Search strategy	Hits retrieved
1. CENTRAL (the Cochrane Library) 1996 to present cr-so.cochrane.org/SearchSimple.php (Date of most recent search: 2 June 2021)	#1 MESH DESCRIPTOR DEMENTIA EXPLODE ALL TREES 5587	April 2020: 462
	#2 MESH DESCRIPTOR DEMENTIA EXPLODE ALL TREES 5587	June 2021: 139
	#3 major cognitive disorder 1	
	#4 alzheimer* 11400	
	#5 dement* 21613	
	#6 (lewy adj2 bod*) or LBD or DLB 409	
	#7 FTLD or frontotemp* 489	
	#8 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 25147	
	#9 MESH DESCRIPTOR Neuropsychological Tests EXPLODE ALL TREES 15861	
	#10 MESH DESCRIPTOR Cognition Disorders EXPLODE ALL TREES WITH QUALIFIERS DI 985	
	#11 (((cognit* or memor* or neuropsychological*) adj3 (assess* or test* or task* or performance* or decline* or function*))) :TI,AB,KY 31618	
	#12 MoCA:TI,AB,KY 838	
	#13 MMSE:TI,AB,KY 3504	
	#14 (Mini-mental State Examination):TI,AB,KY 4195	
	#15 (Brief Screen for Cognition Impairment):TI,AB,KY 0	

(Continued)

- #16 (Memory and Ageing Telephone Screen):TI,AB,KY 2
- #17 (Telephone Cognitive Assessment Battery):TI,AB,KY 1
- #18 (Short Portable Mental Status Questionnaire):TI,AB,KY 38
- #19 (Telephone Modified Mini- Mental state exam):TI,AB,KY 1
- #20 (Telephone administered Minnesota Cognitive Acuity Screen):TI,AB,KY 0
- #21 (Blessed Telephone Information Memory Concentration Test):TI,AB,KY 0
- #22 (Structured telephone interview for dementia assessment):TI,AB,KY 0
- #23 TICS:TI,AB,KY 385
- #24 TICSm:TI,AB,KY 3
- #25 TICS-M:TI,AB,KY 21
- #26 sMMSE:TI,AB,KY 39
- #27 (Telephone Interview for Cognitive Status):TI,AB,KY 31
- #28 #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 44059
- #29 #8 AND #28 10661
- #30 MESH DESCRIPTOR Internet EXPLODE ALL TREES 3750
- #31 MESH DESCRIPTOR Smartphone EXPLODE ALL TREES 312
- #32 MESH DESCRIPTOR Telecommunications EXPLODE ALL TREES 5972
- #33 camera* 2405
- #34 phone* 10749
- #35 Smartphone 2769
- #36 teleconferenc* 217
- #37 telephone* 16506
- #38 telepsychiatry 84
- #39 telemedicine* 3553
- #40 video* 18901
- #41 webcam* 54
- #42 (remote* adj (test* or diagnos* or consult* or deliver*)) 667
- #43 mobile tablet 15
- #44 #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 53114
- #45 #29 AND #44 462

2. MEDLINE In-process and other non-indexed citations and MEDLINE 1950-present (OvidSP)	1 exp DEMENTIA/	April 2020: 1123
	2 major cognitive disorder.ti,ab.	June 2021: 146
	3 alzheimer*.ti,ab.	

(Continued)

(Date of most recent
search: 2 June 2021)

- 4 dement*.ti,ab.
- 5 ((lewy adj2 bod*) or LBD or DLB).ti,ab.
- 6 (FTLD or frontotemp*).ti,ab.
- 7 or/1-6
- 8 exp Neuropsychological Tests/
- 9 exp Cognition Disorders/di [Diagnosis]
- 10 ((cognit* or memor* or neuropsychological*) adj3 (assess* or test* or task* or performance* or decline* or function*)).ti,ab.
- 11 MoCA.ti,ab.
- 12 MMSE.ti,ab.
- 13 "Mini-mental State Examination".ti,ab.
- 14 "Brief Screen for Cognition Impairment".ti,ab.
- 15 "Memory and Ageing Telephone Screen".ti,ab.
- 16 "Telephone Cognitive Assessment Battery".ti,ab.
- 17 "Short Portable Mental Status Questionnaire".ti,ab.
- 18 "Telephone Modified Mini- Mental state exam".ti,ab.
- 19 "Telephone administered Minnesota Cognitive Acuity Screen".ti,ab.
- 20 "Blessed Telephone Information Memory Concentration Test".ti,ab.
- 21 "Structured telephone interview for dementia assessment".ti,ab.
- 22 TICS.ti,ab.
- 23 TICSm.ti,ab.
- 24 TICS-M.ti,ab.
- 25 sMMSE.ti,ab.
- 26 "Telephone Interview for Cognitive Status".ti,ab.
- 27 or/8-26
- 28 7 and 27
- 29 exp Internet/
- 30 Smartphone/
- 31 exp Telecommunications/
- 32 camera*.ti,ab.
- 33 phone*.ti,ab.
- 34 Smartphone.ti,ab.
- 35 teleconferenc*.ti,ab.
- 36 telephone*.ti,ab.
- 37 telepsychiatry.ti,ab.

(Continued)

- 38 telemedicine*.ti,ab.
- 39 video*.ti,ab.
- 40 webcam*.ti,ab.
- 41 (remote* adj (test* or diagnos* or consult* or deliver*)).ti,ab.
- 42 "mobile tablet".ti,ab.
- 43 or/29-42
- 44 28 and 43

3. Embase 1974 to present (OvidSP) (Date of most recent search: 2 June 2021)	1 Dementia/ 2 Delirium/ 3 Wernicke Encephalopathy/ 4 Delirium, Dementia, Amnestic, Cognitive Disorders/ 5 ("benign senescent forgetfulness" or ("normal pressure hydrocephalus" and "shunt*") or ("organic brain disease" or "organic brain syndrome") or ((cerebral* or cerebrovascular or cerebro-vascular) adj2 insufficien*) or (cerebr* adj2 deteriorat*) or (chronic adj2 (cerebrovascular or cerebro-vascular)) or (creutzfeldt or jcd or cjd) or (lewy* adj2 bod*) or (pick* adj2 disease) or alzheimer* or binswanger* or deliri* or dement* or huntington* or korsako*).tw. 6 "major neurocognitive disorder".ti,ab. 7 or/1-6 8 exp neuropsychological test/ 9 ((cognit* or memor* or neuropsychological*) adj3 (assess* or test* or task* or performance* or decline* or function*)).ti,ab. 10 MoCA.ti,ab. 11 MMSE.ti,ab. 12 "Mini-mental State Examination".ti,ab. 13 "Brief Screen for Cognition Impairment".ti,ab. 14 "Memory and Ageing Telephone Screen".ti,ab. 15 "Telephone Cognitive Assessment Battery".ti,ab. 16 "Short Portable Mental Status Questionnaire".ti,ab. 17 "Telephone Modified Mini- Mental state exam".ti,ab. 18 "Telephone administered Minnesota Cognitive Acuity Screen".ti,ab. 19 "Blessed Telephone Information Memory Concentration Test".ti,ab. 20 "Structured telephone interview for dementia assessment".ti,ab. 21 TICS.ti,ab. 22 TICSm.ti,ab. 23 TICS-M.ti,ab.	April 2020: 2957 June 2021: 587
--	--	--

(Continued)

- 24 sMMSE.ti,ab.
- 25 "Telephone Interview for Cognitive Status".ti,ab.
- 26 or/8-25
- 27 7 and 26
- 28 exp Internet/
- 29 exp smartphone/
- 30 exp telecommunication/
- 31 camera*.ti,ab.
- 32 phone*.ti,ab.
- 33 Smartphone.ti,ab.
- 34 teleconferenc*.ti,ab.
- 35 telephone*.ti,ab.
- 36 telepsychiatry.ti,ab.
- 37 telemedicine*.ti,ab.
- 38 video*.ti,ab.
- 39 webcam*.ti,ab.
- 40 (remote* adj (test* or diagnos* or consult* or deliver*)).ti,ab.
- 41 "mobile tablet".ti,ab.
- 42 or/28-41
- 43 27 and 42

4. PSYCINFO (OvidSP)	1 exp Dementia/	April 2020: 817
1927 to present	2 exp Delirium/	June 2021: 104
(Date of most recent search: 2 June 2021)	3 exp Huntingtons Disease/	
	4 exp Kluver Bucy Syndrome/	
	5 exp Wernickes Syndrome/	
	6 exp Cognitive Impairment/	
	7 dement*.mp.	
	8 alzheimer*.mp.	
	9 (lewy* adj2 bod*).mp.	
	10 deliri*.mp.	
	11 (chronic adj2 cerebrovascular).mp.	
	12 ("organic brain disease" or "organic brain syndrome").mp.	
	13 "supranuclear palsy".mp.	
	14 ("normal pressure hydrocephalus" and "shunt").mp.	

(Continued)

- 15 "benign senescent forgetfulness".mp.
- 16 (cerebr* adj2 deteriorat*).mp.
- 17 (cerebral* adj2 insufficient*).mp.
- 18 (pick* adj2 disease).mp.
- 19 (creutzfeldt or jcd or cjd).mp.
- 20 huntington*.mp.
- 21 binswanger*.mp.
- 22 korsako*.mp.
- 23 ("parkinson* disease dementia" or PDD or "parkinson* dementia").mp.
- 24 "major neurocognitive disorder".ti,ab.
- 25 or/1-24
- 26 exp Neuropsychological Assessment/
- 27 ((cognit* or memor* or neuropsychological*) adj3 (assess* or test* or task* or performance* or decline* or function*)).ti,ab.
- 28 MoCA.ti,ab.
- 29 MMSE.ti,ab.
- 30 "Mini-mental State Examination".ti,ab.
- 31 "Brief Screen for Cognition Impairment".ti,ab.
- 32 "Memory and Ageing Telephone Screen".ti,ab.
- 33 "Telephone Cognitive Assessment Battery".ti,ab.
- 34 "Short Portable Mental Status Questionnaire".ti,ab.
- 35 "Telephone Modified Mini- Mental state exam".ti,ab.
- 36 "Telephone administered Minnesota Cognitive Acuity Screen".ti,ab.
- 37 "Blessed Telephone Information Memory Concentration Test".ti,ab.
- 38 "Structured telephone interview for dementia assessment".ti,ab.
- 39 TICS.ti,ab.
- 40 TICSm.ti,ab.
- 41 TICS-M.ti,ab.
- 42 sMMSE.ti,ab.
- 43 "Telephone Interview for Cognitive Status".ti,ab.
- 44 or/26-43
- 45 25 and 44
- 46 exp Internet/
- 47 exp Smartphones/
- 48 exp Telecommunications Media/

(Continued)

49 exp Telemedicine/
 50 camera*.ti,ab.
 51 phone*.ti,ab.
 52 Smartphone.ti,ab.
 53 teleconferenc*.ti,ab.
 54 telephone*.ti,ab.
 55 telepsychiatry.ti,ab.
 56 telemedicine*.ti,ab.
 57 video*.ti,ab.
 58 webcam*.ti,ab.
 59 (remote* adj (test* or diagnos* or consult* or deliver*)).ti,ab.
 60 "mobile tablet".ti,ab.
 61 or/46-60
 62 45 and 61

5. CINAHL (EBSCOhost)	S58 S42 AND S57	April 2020: 1234
1984 to present (Date of most recent search: 2 June 2021)	S57 S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 S56 TX mobile tablet S55 TX (remote* n (test* or diagnos* or consult* or deliver*)) S54 TX webcam* S53 TX video* S52 TX telemedicine* S51 TX telepsychiatry S50 TX telephone* S49 TX teleconferenc* S48 TX Smartphone S47 TX phone* S46 TX camera* S45 (MH "Telecommunications+") S44 (MH "Smartphone") S43 (MH "Internet+") S42 S21 AND S41 S41 S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 S40 TX "Telephone Interview for Cognitive Status"	June 2021: 88

(Continued)

S39 TX sMMSE

S38 TX TICS-M

S37 TX TICSm

S36 TX TICS

S35 TX "Structured telephone interview for dementia assessment"

S34 TX "Blessed Telephone Information Memory Concentration Test"

S33 TX "Telephone administered Minnesota Cognitive Acuity Screen"

S32 TX "Telephone Modified Mini- Mental state exam"

S31 TX "Short Portable Mental Status Questionnaire"

S30 TX "Telephone Cognitive Assessment Battery"

S29 TX "Memory and Ageing Telephone Screen"

S28 TX "Brief Screen for Cognition Impairment"

S27 TX "Mini-mental State Examination"

S26 TX MMSE

S25 TX MoCA

S24 TX ((cognit* or memor* or neuropsychological*) n3 (assess* or test* or task* or performance* or decline* or function*))

S23 (MH "Cognition Disorders+/DI")

S22 (MH "Neuropsychological Tests+")

S21 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20

S20 TX "major neurocognitive disorder"

S19 TX korsako*

S18 TX binswanger*

S17 TX huntington*

S16 TX creutzfeldt or jcd or cjd

S15 TX pick* N2 disease

S14 TX cerebral* N2 insufficient*

S13 TX cerebr* N2 deteriorat*

S12 TX "benign senescent forgetfulness"

S11 TX "normal pressure hydrocephalus" and "shunt**"

S10 TX "organic brain disease" or "organic brain syndrome"

S9 TX chronic N2 cerebrovascular

S8 TX deliri*

S7 TX lewy* N2 bod*

(Continued)

S6 TX alzheimer*

S5 TX dement*

S4 MH "Wernicke's Encephalopathy

S3 MH "Delirium, Dementia, Amnestic, Cognitive Disorders"

S2 MH "Delirium"

S1 (MH "Dementia+")

<p>6. Web of Science (Clarivate) – 1900 to present.</p> <p>Science Citation Index Expanded (SCIE); Social Sciences Citation Index (SSCI); Arts & Humanities Citation Index (AHCI); Emerging Sources Citation Index (ESCI); Conference Proceedings Citation Index (CPCI); Book Citation Index (BKCI); Current Chemical Reactions and Index Chemicus</p> <p>(Date of most recent search: 2 June 2021)</p>	<p>TOPIC: (dement* OR alzheimer* OR "vascular cognitive impairment" OR "leu* bod*" OR CADASIL OR "cognit* impair*" OR FTD OF FTLD OR "cerebrovascular insufficienc*" OR AD OR VCI) AND TOPIC: (Neuropsychological Test* OR MoCA OR MMSE OR COGNITIVE TEST* OR MEMORY TEST* OR NEUROPSYCHOLOGICAL TEST*) AND TOPIC: (Internet OR Smartphone OR Telecommunications OR CAMERA* OR PHONE* OR TELEPHONE* OR telemedicine OR VIDEO* OR WEB-CAM*)</p>	<p>April 2020: 1502</p> <p>June 2021: 266</p>
<p>7. LILACS (BIREME)</p> <p>1978 to present</p> <p>accessed lilacs.bvsalud.org/en/</p> <p>(Date of most recent search: 2 June 2021)</p>	<p>alzheimer OR alzheimers OR alzheimer's OR dementia OR demenc\$ [Words] and Internet OR Smartphone OR Telecommunications OR camera OR phone OR telephone OR video OR webcam [Words] and MoCA OR MMSE OR COGNITIVE TEST* OR MEMORY TEST* OR NEUROPSYCHOLOGICAL TEST* OR Diagnos* [Words]</p>	<p>April 2020: 0</p> <p>June 2021: 0</p>
<p>8. ClinicalTrials.gov</p> <p>2000 to present</p> <p>(www.clinicaltrials.gov)</p> <p>(Date of most recent search: 2 June 2021)</p>	<p>Internet OR Smartphone OR Telecommunications OR camera OR phone OR telephone OR video OR webcam dementia OR alzheimers OR cognition OR cognitive MoCA OR MMSE OR COGNITIVE TEST* OR MEMORY TEST* OR NEUROPSYCHOLOGICAL TEST* OR Diagnos*</p>	<p>April 2020: 103</p> <p>June 2021: 5</p>
<p>9. ALOIS CDCIG specialised register (CRS web)</p> <p>2008 to present</p> <p>Accessed crsweb.cochrane.org/login.html</p> <p>(Date of most recent search: 2 June 2021)</p>	<p>#1 Internet OR Smartphone OR Telecommunications OR CAMERA* OR PHONE* OR TELEPHONE* OR telemedicine OR VIDEO* OR WEBCAM AND INREGISTER</p> <p>#2 Neuropsychological Test* OR MoCA OR MMSE OR COGNITIVE TEST* OR MEMORY TEST* AND INREGISTER</p> <p>#3 #1 AND #2</p>	<p>April 2020: 258</p> <p>June 2021: 55</p>

(Continued)

TOTAL before deduplication	April 2020:8456
	June 2021: 1390
	Total: 9846
TOTAL after de-duplication	April 2020: 5600
	June 2021: 990
	Total: 6590

Appendix 2. QUADAS-2 anchoring statements

We provide some core anchoring statements for quality assessment of diagnostic test accuracy reviews of neuropsychological tests in dementia. These statements are designed for use with the QUADAS-2 tool and were derived during a two-day, multidisciplinary focus group in 2010. If a QUADAS-2 signalling question for a specific domain is answered 'yes', the risk of bias can be judged as 'low'. If a question is answered 'no', this indicates a risk of potential bias. The focus group was tasked with judging the extent of the bias for each domain. During this process, it became clear that certain issues were key to assessing quality, while others were important to record but were less important for assessing overall quality. To assist, we describe a 'weighting' system. When an item is weighted 'high risk', that section of the QUADAS-2 results table is judged to have a high potential for bias if a signalling question is answered 'no'. For example, in dementia diagnostic test accuracy studies, ensuring that clinicians performing dementia assessment are blinded to results of the index test is fundamental. If this blinding was not present, the item on the reference standard should be scored 'high risk of bias', regardless of the other contributory elements. When an item is weighted 'low risk', it is judged to have a low potential for bias if a signalling question for that section of the QUADAS-2 results table is answered 'no'. Overall bias will be judged on whether other signalling questions (with a high risk of bias) for the same domain are also answered 'no'. In assessing individual items, a score of 'unclear' should be given only if there is genuine uncertainty. In these situations, the review authors will contact the relevant study teams for additional information.

Anchoring statements to assist with risk of bias assessment

Domain 1: patient selection

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

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

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

Did the study avoid inappropriate exclusions? The study was automatically graded as unclear if exclusions were not detailed (pending contact with study authors). When exclusions were detailed, we graded the study as 'low risk' if we considered that the exclusions were appropriate. Certain exclusions common to many studies of dementia are medical instability, terminal disease, alcohol/substance misuse, concomitant psychiatric diagnosis and other neurodegenerative condition. However, if 'difficult to diagnose' groups are excluded, this may introduce bias, so exclusion criteria must be justified. For a community sample, we would expect relatively few exclusions. We labelled post hoc exclusions as 'high risk' of bias. Weighting: high risk of bias.

Applicability: are there concerns that the included participants do not match the review question? (high/low/unclear)

The included participants should match the intended population as described in the review question. If not already specified in the review inclusion criteria, the setting will be particularly important – the review authors should consider population in terms of symptoms, pretesting and potential disease prevalence. We classified studies that used very selected participants or subgroups as low applicability, unless they were intended to represent a defined target population, for example, people with memory problems referred to a specialist and investigated by lumbar puncture.

Domain 2: index test

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

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

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

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

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

Variations in the length, structure, language, or administration (or a combination of these) of the index test may all affect applicability if they differed from those specified in the review question.

Domain 3: reference standard

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

Is the reference standard likely to correctly classify the target condition? Commonly used international criteria that can assist with clinical diagnosis of dementia include those detailed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and the International Classification of Diseases (ICD-10). Criteria specific to dementia subtypes include, but are not limited to, the National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for Alzheimer's dementia; McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; and National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria for vascular dementia. When the criteria used for assessment were unfamiliar to the review authors and the Cochrane Dementia and Cognitive Improvement Group, this item was classified as 'high risk of bias'. Weighting: high risk.

Were the reference standard results interpreted without knowledge of the results of the index test? Terms such as 'blinded' or 'independent' were sufficient; full details of the blinding procedure were not required. This may have been scored as 'low risk' if explicitly described, or if a clear temporal pattern to the order of testing was evident (e.g. all dementia assessments performed before (neuropsychological test) testing). Informant rating scales and direct cognitive tests present certain problems. It is accepted that informant interview and cognitive testing are usual components of clinical assessment for dementia; however, specific use of the scale under review in the clinical dementia assessment should be scored as high risk of bias. Weighting: high risk.

Was sufficient information on the method of dementia assessment given for the assessment to be repeated in an independent study? Particular points of interest for dementia assessment include the training/expertise of the assessor; whether additional information (e.g. neuroimaging; other neuropsychological test results) was available to inform the diagnosis and whether this was available for all participants. Weighting: variable risk, but high risk if method of dementia assessment not described.

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

There exists the possibility that some methods of dementia assessment, although valid, may diagnose a smaller or larger proportion of participants with disease than in usual clinical practice. In these instances, the item should be rated 'poor applicability'.

Domain 4: patient flow and timing (note: refer to, or construct, a flow diagram)

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

Was there an appropriate interval between the index test and the reference standard? For a cross-sectional study design, the potential exists for the participant to change between assessments; however, dementia is a slowly progressive disease that is not reversible. The ideal scenario would be a same-day assessment, but longer periods of time (e.g. several weeks or months) are unlikely to lead to a high risk

of bias. For delayed-verification studies, the index and reference tests are necessarily separated in time, given the nature of the condition. Weighting: low risk.

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

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

HISTORY

Protocol first published: Issue 9, 2020

CONTRIBUTIONS OF AUTHORS

TQ conceived the idea.

All authors contributed to writing the protocol and responding to peer review comments.

The Cochrane Dementia and Cognitive Improvement Group Information Specialists designed the search.

DECLARATIONS OF INTEREST

LB: none.

EE: none.

TH: none.

RM: none.

AO: none.

AE: none.

TQ: none.

SOURCES OF SUPPORT

Internal sources

- No sources of support provided

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 therein 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

- Dunhill Medical Trust, UK

Salary for Dr Lucy Beishon.

- NIHR, UK

NIHR Academic Clinical Lecturer supporting the salary for Dr Lucy Beishon.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We did not perform a quantitative meta-analysis due to insufficient studies reporting test accuracy data for the same test at the same threshold. In the protocol, we planned to exclude studies where the interval between the index test and reference standard was longer than one month. Due to a lack of available studies we have included studies with a longer test interval but have discussed this as a potential

source of heterogeneity among the included studies in the findings section. We excluded studies where the same assessor conducted the in-person and remote assessments on the same day due to the high risk of bias.

G Martínez and Z Tieges left the review author team and Lucy C Beishon, Amy R Elliott, and Aoife O'Mahony joined the team.

INDEX TERMS

Medical Subject Headings (MeSH)

Cognition; Cross-Sectional Studies; *Dementia [diagnosis]; Diagnostic Tests, Routine; Reproducibility of Results; Sensitivity and Specificity; Telephone

MeSH check words

Humans