Original Contribution

Test Accuracy of Cognitive Screening Tests for Diagnosis of Dementia and Multidomain Cognitive Impairment in Stroke

Rosalind Lees, MA; Johann Selvarajah, PhD; Candida Fenton, MSc; Sarah T. Pendlebury, DPhil; Peter Langhorne, PhD; David J. Stott, MD; Terence J. Quinn, MD

Background and Purpose—Guidelines recommend screening stroke-survivors for cognitive impairments. We sought to collate published data on test accuracy of cognitive screening tools.

Methods—Index test was any direct, cognitive screening assessment compared against reference standard diagnosis of (undifferentiated) multidomain cognitive impairment/dementia. We used a sensitive search statement to search multiple, cross-disciplinary databases from inception to January 2014. Titles, abstracts, and articles were screened by independent researchers. We described risk of bias using Quality Assessment of Diagnostic Accuracy Studies tool and reporting quality using Standards for Reporting of Diagnostic Accuracy guidance. Where data allowed, we pooled test accuracy using bivariate methods.

Results—From 19,182 titles, we reviewed 241 articles, 35 suitable for inclusion. There was substantial heterogeneity:
25 differing screening tests; differing stroke settings (acute stroke, n=11 articles), and reference standards used (neuropsychological battery, n=21 articles). One article was graded low risk of bias; common issues were case–control methodology (n=7 articles) and missing data (n=22). We pooled data for 4 tests at various screen positive thresholds: Addenbrooke’s Cognitive Examination-Revised (<88/100): sensitivity 0.96, specificity 0.70 (2 studies); Mini Mental State Examination (<27/30): sensitivity 0.71, specificity 0.85 (12 studies); Montreal Cognitive Assessment (<26/30): sensitivity 0.95, specificity 0.45 (4 studies); MoCA (<22/30): sensitivity 0.84, specificity 0.78 (6 studies); Rotterdam-CAMCOG (<33/49): sensitivity 0.57, specificity 0.92 (2 studies).

Conclusions—Commonly used cognitive screening tools have similar accuracy for detection of dementia/multidomain impairment with no clearly superior test and no evidence that screening tools with longer administration times perform better. MoCA at usual threshold offers short assessment time with high sensitivity but at cost of specificity; adapted cutoffs have improved specificity without sacrificing sensitivity. Our results must be interpreted in the context of modest study numbers: heterogeneity and potential bias. (Stroke. 2014;45:00-00.)

Key Words: cognitive impairment ■ dementia ■ sensitivity ■ specificity ■ stroke

Stroke-survivors are at particular risk of cognitive decline. Three month dementia prevalence is ≥30%, and even minor stroke events have cognitive sequel.1,2 Poststroke cognitive impairment is associated with increased mortality, disability, and institutionalization.3 The importance of cognitive change is highlighted by stroke-survivors themselves. In a national priority setting exercise, cognitive impairment was voted the single most important topic for stroke research.4 A first step in management of cognitive problems is recognition and diagnosis. Informal clinician assessment will miss important cognitive problems,5 and formal cognitive testing is recommended.6–8 The ideal would be expert, multidisciplinary assessment informed by comprehensive investigations. This approach is not feasible at a population level. In practice, a 2-step system is adopted, with baseline cognitive testing used for screening or triage and specialist assessment to define the cognitive problem offered depending on the results.

Although there is general agreement on the merits of poststroke cognitive assessment, there is no consensus on a preferred testing strategy.5,6 Various cognitive screening tools are available with substantial variation in test used.9–10

The clinical meaning of cognitive problems after stroke will vary according to test context. Cognitive impairment diagnosed in the first days post stroke may reflect a mix of delirium, stroke-specific impairments, and prestroke cognitive decline.2,11,12 In the longer term, assessments aim to make or refute a dementia diagnosis. Common to all test situations is a final diagnosis of presence/absence of clinically important impairments. A screening assessment should detect this clinical syndrome of all-cause, poststroke multidomain cognitive impairment.

Received April 16, 2014; accepted July 31, 2014.

From the Institute of Cardiovascular and Medical Sciences (R.L., P.L., D.J.S., T.J.Q.), Institute of Neuroscience and Psychology (J.S.), MRC/CSO Social and Public Health Sciences Unit (C.F.), University of Glasgow, UK; and NIHR Biomedical Research Centre and Stroke Prevention Research Unit, John Radcliffe Hospital, Oxford, UK (S.T.P.).

Presented in part at the European Stroke Conference, Nice, France, May 6–9, 2014.

The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.114.005842/-/DC1.

Correspondence to Terence J. Quinn, MD, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Room 2.44, New Lister Building, Glasgow Royal Infirmary, 10-16 Alexandra Parade, Glasgow, UK. E-mail terry.quinn@glasgow.ac.uk

© 2014 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org

DOI: 10.1161/STROKEAHA.114.005842
Collation and synthesis of the evidence around cognitive screening tools will help guide policy and practice and highlight knowledge gaps. We sought to perform systematic review and meta-analysis, describing accuracy of screening tools for assessing dementia and multidomain cognitive impairment in stroke-survivors.

**Methods**

We used systematic literature review and meta-analysis techniques specific to test accuracy and followed best practice in reporting.

**Aims**

Coprimary aims were to describe the following:

1. Accuracy of cognitive screening tests for clinical diagnosis of multidomain, cognitive impairment/dementia in stroke-survivors.
2. Accuracy of brief (<5 minutes completion time), cognitive screening tests against more detailed neuropsychological assessments.

If data allowed, secondary objectives were to compare differing cut-off scores used to define a threshold of test positivity and to compare the effects of heterogeneity with specific reference to test context and diagnostic reference standard.

**Index Test**

Index tests of interest were any direct-to-patient, cognitive screening tests. We included any screening test where the authors described it as such. We excluded informant-based assessments and tests that require testing equipment considered nonstandard for a stroke service. We recognize that language and visuospatial function are important components of cognition, but did not include assessments of tools designed to exclusively test these domains. We did not include studies that compared screening tools with no reference to diagnostic gold standard.

**Target Condition and Reference Standard**

Our target condition of interest was all-cause (undifferentiated) multidomain cognitive impairment post stroke. This rubric recognizes that a diagnosis of important poststroke cognitive problems can be made without necessarily assigning a dementia label or specifying a pathological subtype. As reference standard, we included clinical diagnosis of dementia made using any recognized classification system. We also included multidomain, cognitive problems as detailed on a neuropsychological assessment, provided the test battery was comprehensive and the results were interpreted to give a diagnostic formulation. We did not include single domain cognitive impairment or cognitive impairment no dementia because of inconsistency in operationalization of these syndromes. We prespecified subgroup analyses comparing dementia diagnosis and neuropsychological battery-based diagnosis.

For assessment of brief tests, we accepted results from a more detailed multidomain, screening assessment as reference standard, for example, comparison of Hodgkinson’s abbreviated mental test against Folstein’s Mini Mental State Examination (MMSE).

**Participants and Setting**

Our focus was stroke-survivors. Where study population was a mix of stroke survivors and other patients, we included if proportion of stroke-survivors was >75%. We made no distinction between stroke subtypes, but excluded studies of traumatic intracerebral hemorrhage and subarachnoid hemorrhage. We did not include case studies (defined as having fewer than 10 participants).

We included studies conducted in any clinical setting and at any time post stroke. We operationalized time since stroke as hyperacute (first 7 days); acute (8–14 days); post acute (15 days to 3 months); medium term (3–12 months), and longer term (post 1 year).

**Search Strategy**

All aspects of searching, data extraction, and study assessment were performed by 2 reviewers (RL, JS) based in different centers and blinded to each other’s results. On review of paired data, disagreement was resolved by group discussion.

We developed a sensitive search strategy in collaboration with an Information Scientist (CF) and with assistance from the Cochrane Dementia and Cognitive Improvement Group. Search terms were developed using a concepts-based approach, using Medical Subject Heading terms and other controlled vocabulary. Concepts of interest were stroke, dementia, and cognitive assessment. Our cognitive assessment concept included terms relating to cognitive tests used in stroke, based on previous survey data. We searched multiple, international, cross-disciplinary electronic databases from inception to January 2014. (Full search strategy is detailed in the online-only Data Supplement I–II.) We checked reference lists of relevant studies and reviews for further titles, repeating the process until no new titles were found.

We screened all titles generated by initial searches for relevance, and corresponding abstracts were assessed and potentially eligible studies reviewed as full manuscripts against inclusion criteria. As a check of internal validity, a random selection of 2000 titles from the original search was reassessed by a third author (TQ). We tested external validity of our search by sharing lists of included articles with independent researchers (Acknowledgments). One author (TQ) identified exemplar articles relevant to the review question (online-only Data Supplement II), and we assessed whether these titles were detected by search strategy.

**Data Extraction and Management**

We extracted data to a study-specific proforma piloted against 2 articles (available from contact author).

For screening tests that give an ordinal summary score, various cut-points can be used to define test positive cases. Where data were given for several thresholds, we extracted separate data for each. Where a study may have included usable data, but these were not presented in the published article, we contacted the authors directly. If the same data set was presented in >1 publication, we included the primary article.

**Quality Assessment**

We assessed quality of study reporting using the dementia-specific extension to the Standards for Reporting of Diagnostic Accuracy checklist. We assessed methodological quality using the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2). We have previously developed QUADAS-2 anchoring statements specific to cognitive assessment.

**Statistical Analysis**

We assessed accuracy of screening tests against a dichotomous variable cognitive impairment/no cognitive impairment. We created standard 2-by-2 data tables describing binary test results cross-classified with binary reference standard. We did not include case–control studies in pooled analyses. We used accepted cut-offs for multidomain impairment/dementia published in the literature: Addenbrooke’s Cognitive Examination-Revised <88, Rotterdam-CAMCOG<33, MMSE<27, and <25. For Montreal Cognitive Assessment (MoCA), a cut-off of <26 was used together with the lower cut-off of <22 because the former was developed to detect single-domain impairment. Where data allowed, we calculated sensitivity, specificity, and corresponding 95% confidence intervals (95% CI) and created test accuracy forest plots (RevMan 5.1, Cochrane Collaboration). We pooled test accuracy data using the bivariate approach (Statistical Analysis Software v9.1; SAS Institute Inc. USA). We used a bespoke macro developed with assistance of a statistical team with an interest in test accuracy. Summary metrics of interest were sensitivity/specificity and positive/negative likelihood ratios and clinical utility index. We created summary curves in received operating characteristic space with corresponding 95% prediction intervals.

We assessed potential heterogeneity through visual inspection of forest plots. We prespecified 2 factors that may contribute to heterogeneity, timing of assessment and reference standard. We dichotomized studies into acute (classified as hyperacute or acute) or nonacute and described reference standard as clinical (clinical diagnosis of dementia) and...
neuropsychological battery (multidomain cognitive impairment). We assessed effect by plotting summary received operating characteristic curves by covariate. Taking timing of assessment as example, we calculated sensitivity/specificity restricted to the single cognitive assessment with greatest number of studies. We described a relative sensitivity/specificity comparing acute and nonacute studies, where a result of unity suggests no difference in that diagnostic property between settings. We did not quantify publication bias because there is no assessment applicable to test accuracy.

Where articles fulfilled inclusion criteria but did not have data suitable for this method of analysis, we offer a tabulated/narrative description.

Results

From 19,182 titles, we reviewed 241 full articles of which 35 (34 data sets) were suitable for inclusion. Scope of included literature was international, with articles from 16 different countries.

We detailed the study selection process in a Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flow diagram (Figure 1).

Our validation check suggested the initial search was appropriate because all prespecified articles were found on first search.

---

**Figure 1.** Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flow diagram. ACE-R indicates Addenbrooke’s cognitive examination-revised; AMT, Abbreviated Mental Test; FN, false-negative; FP, false-positive; HSROC, Hierarchical Summary ROC; MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; R-CAMCOG, Rotterdam-CAMCOG; ROC, received operating characteristic; TN, true-negative; and TP, true-positive.
Accuracy of Screening Tools for Diagnosis of Cognitive Impairment

In total n=32 articles (n=3562 participants) were eligible.21–53 We tabulated summary descriptors for studies using clinical diagnosis reference standard (n=11 articles)22,24,34,37–39,43,49,51,52 and those using detailed neuropsychological assessment (n=21; Tables 1–2).21,23,25-33,35,36,40–42,44–46,48,50

There was considerable heterogeneity in study population: setting and test strategy. Twenty-three different tests were described, commonest MMSE (n=16 articles) and MoCA (n=8; Tables 1–2; online-only Data Supplement III). Screening tests gave a spread of test accuracy, sensitivity range 14% to 100%, and specificity range 0% to 100% with trade-off between sensitivity and specificity. Where authors compared >1 test in the same population, the comparator was usually MMSE, and for most articles, MMSE was more specific and less sensitive than other tests. Both the Executive Function Performance Test and the Repeatable Battery for Assessment of Neuropsychological Status showed strong correlations with scores on neuropsychological batteries, but data were not suitable for pooled analysis.25–28

Quality and Reporting

One article was graded low risk for all QUADAS2 domains.21

Common issues of concern were use of case–control methodology (n=7 articles) and potential lack of blinding (n=12; Figure 2; online-only Data Supplement IV). Five articles attempted to include patients with moderate to severe aphasia.21–25

Standards for Reporting of Diagnostic Accuracy assessment suggested consistent areas of poor reporting, particularly around the handling of missing data (n=22 articles) and descriptions of training and expertise of assessors (n=25 articles; online-only Data Supplement V).

Meta-Analyses

We were able to pool test accuracy data for 4 tests: Addenbrooke’s Cognitive Examination-Revised (threshold score ≤88); MMSE (thresholds ≤24 and ≤26), MoCA (thresholds <26 and <22), and Rotterdam-CAMCOG (threshold <33). No test had sensitivity and specificity that were significantly different from others. MoCA at traditional threshold was sensitive at cost of specificity; specificity improved if thresholds were adjusted (Table 3; Figure 3; online-only Data Supplement VI).

Heterogeneity

We assessed effect of timing and reference standard using MMSE data. Comparing 6 acute studies20,21,31,37,38,45 and 6 nonacute studies22,30,34,39,46,50, we gave a relative sensitivity, 0.73 (95% CI, 0.58–0.93) and relative specificity, 1.12 (95% CI, 1.01–1.25). These data suggest that accuracy varies with assessment timing and context, with acute testing yielding higher sensitivity and lower specificity. From analysis of paired summary received operating characteristic curves, the overall accuracy was better for those studies where assessment was performed within the acute period (online-only Data Supplement VII). Comparing clinical dementia reference standard21,24,40,47,52 against neuropsychological battery23,32,33,36,39,41,45,46, suggested no difference in test properties dependent on the reference standard used, with relative sensitivity of 0.86 (95% CI, 0.67–1.11) and relative specificity of 1.05 (95% CI, 0.95–1.16).

<table>
<thead>
<tr>
<th>Table 1. Test Accuracy Data for Studies Where Reference Standard was Clinical Diagnosis of Dementia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Brookes34*</td>
</tr>
<tr>
<td>Cumming 201022</td>
</tr>
<tr>
<td>de Koning 200032</td>
</tr>
<tr>
<td>de Koning 200538</td>
</tr>
<tr>
<td>de Koning52</td>
</tr>
<tr>
<td>Dong 201223</td>
</tr>
<tr>
<td>Hershey42</td>
</tr>
<tr>
<td>Hobson42*</td>
</tr>
<tr>
<td>Srikanth24</td>
</tr>
<tr>
<td>Tang47</td>
</tr>
<tr>
<td>Wu49</td>
</tr>
</tbody>
</table>

Where >1 test threshold was described, we present the primary data. AD indicates Alzheimer’s disease; BMET, brief memory and executive test; CSCSE, cognitive capacity screening examination; MDRS, Mattis dementia rating scale; MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; N/A, not applicable; PNB, preliminary neuropsychological battery; R-CAMCOG, Rotterdam-CAMCOG; and SVD, small vessel disease.

*Case–control study, data not included in pooled analyses.
Accuracy of Brief Tools

We found 3 suitable articles (n=294 participants; online-only Data Supplement VIII).53–55 Two articles were graded high risk of bias and applicability concerns because of case–control methodology and patient inclusion.53,55

Discussion

Previous systematic reviews of cognitive test accuracy have focused on older adults with narrative results only. Our aim was to provide a stroke-specific, literature synthesis to allow evidence-based recommendations on cognitive testing. We were partially successful in this aim. Although there is an extensive, cross-disciplinary literature describing cognitive testing in stroke, number of articles using the classical test accuracy paradigm of index test versus reference (gold) standard was limited. Eligible articles were characterized by substantial heterogeneity and risk of bias, and the potential to describe summary analyses at individual test level was limited.

Accepting these caveats, we can still offer conclusions from our pooled analysis. There was no clearly superior cognitive screening test. Given the relative consistency in accuracy, choice of test strategy should be informed by other factors,
such as purpose of testing, feasibility; acceptability, and opportunity cost. Recent guidance and practice has tended to favor novel tests compared with the traditionally popular MMSE.6,7,36 Although there may be good reasons to favor other tests, our data do not suggest that MMSE is inferior for the diagnosis of multidomain impairment, albeit MMSE may lack sensitivity for single domain impairment.15,46,56 There was a trend toward better clinical utility for Rotterdam-CAMCOG, but the small number of studies and corresponding wide confidence intervals precludes definitive recommendation.

Preferred test properties will depend on test purpose, particularly the level of cognitive impairment to be detected. It could be argued that for initial screening, sensitivity is preferred compared with specificity. In this case, MoCA and Addenbrooke’s Cognitive Examination-Revised may be preferred compared with specificity. In this case, MoCA and Addenbrooke’s Cognitive Examination-Revised may be preferred. The high false-positive rate for multidomain impairment obtained using MoCA <26 is to be expected because this threshold was chosen for detection of single domain impairment/mild cognitive impairment.15,46 Our findings suggest that an adjusted cut-off <22 has improved overall test properties for multidomain impairment in stroke as suggested previously.15,46

We were pragmatic in our choice of index test and reference standard. Our screening test rubric included relatively short assessments (MoCA) through to fairly lengthy batteries (Repeatable Battery for Assessment of Neuropsychological Status). We did not find significant improvements in sensitivity/specificity comparing shorter and longer screens (MoCA and Addenbrooke’s Cognitive Examination-Revised). This may suggest that increasing the length of test batteries does not necessarily improve the test accuracy.

There is no universally accepted gold standard for dementia. Rather than restricting to a particular pathological subtype, our focus was all cause dementia post stroke. We think this approach is in keeping with current clinical practice, where initial screening highlights potential cognitive issues that can subsequently be characterized with more detailed assessments. Literature describing screening tests for diagnosis of specific dementia subtypes was limited, and we were only able to describe test accuracy for all cause (undifferentiated) dementia. A priori, we decided to include multidomain impairment based on neuropsychological assessment in our reference standard, and this approach maximized potential for pooled analysis and recognizes that clinical diagnosis in certain stroke situations is not always feasible or appropriate. We used multidomain rather than single domain impairment because this is closest to current operational classifications of dementia. We did not specify content of the neuropsychological assessment and recognize the substantial heterogeneity in batteries used.15

There were many potential sources of heterogeneity (sample size, case–control methodology, investigator training); indeed a degree of heterogeneity is expected in test accuracy meta-analyses. We were not able to assess all potentially contributory factors across the modest number of included studies and accept that this limits our findings. We chose to focus on reference standard and timing of assessment as the most important possible sources of heterogeneity. We found that test properties varied with timing and seemed to favor earlier assessment, albeit our rubric of acute assessment included studies from first hours ≤14 days post ictus. Comparative analysis of test properties require careful interpretation; a conservative interpretation would be

Table 3. Pooled Test Accuracy for 4 Cognitive Screening Tests Against a Reference Standard of Cognitive Impairment

<table>
<thead>
<tr>
<th>Test (Threshold)</th>
<th>Articles (Patients)</th>
<th>Cognitive Impairment, n (%)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive Likelihood Ratio (95% CI)</th>
<th>Negative Likelihood Ratio (95% CI)</th>
<th>Clinical Utility Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE-R (&lt;38/100)</td>
<td>2 (192)</td>
<td>52 (27%)</td>
<td>0.96 (0.90–1.0)</td>
<td>0.70 (0.59–0.80)</td>
<td>3.19 (2.24–4.54)</td>
<td>0.06 (0.01–0.22)</td>
<td>+0.67</td>
</tr>
<tr>
<td>MMSE (&lt;25/30)</td>
<td>12 (1639)</td>
<td>483 (30%)</td>
<td>0.71 (0.60–0.80)</td>
<td>0.85 (0.80–0.89)</td>
<td>4.73 (3.63–6.17)</td>
<td>0.34 (0.25–0.47)</td>
<td>+0.46</td>
</tr>
<tr>
<td>MMSE (&lt;27/30)</td>
<td>5 (445)</td>
<td>195 (44%)</td>
<td>0.88 (0.82–0.92)</td>
<td>0.62 (0.50–0.73)</td>
<td>2.33 (1.72–3.17)</td>
<td>0.19 (0.13–0.29)</td>
<td>+0.65</td>
</tr>
<tr>
<td>MoCA (&lt;22/30)</td>
<td>6 (726)</td>
<td>289 (39%)</td>
<td>0.84 (0.76–0.89)</td>
<td>0.78 (0.69–0.84)</td>
<td>3.75 (2.77–5.08)</td>
<td>0.20 (0.15–0.29)</td>
<td>+0.59</td>
</tr>
<tr>
<td>MoCA (&lt;26/30)</td>
<td>4 (326)</td>
<td>131 (40%)</td>
<td>0.95 (0.89–0.98)</td>
<td>0.45 (0.34–0.57)</td>
<td>1.73 (1.43–2.10)</td>
<td>0.10 (0.04–0.23)</td>
<td>+0.60</td>
</tr>
<tr>
<td>R-CAMCOG (&lt;33/49)</td>
<td>2 (421)</td>
<td>90 (21%)</td>
<td>0.81 (0.57–0.93)</td>
<td>0.92 (0.87–0.95)</td>
<td>10.18 (6.41–16.18)</td>
<td>0.20 (0.07–0.52)</td>
<td>+0.59</td>
</tr>
</tbody>
</table>

For clinical utility index, +, positive; −, negative. ACE-R indicates Addenbrooke’s cognitive examination-revised; MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; and R-CAMCOG, Rotterdam-CAMCOG.
Figure 3. Summary received operating characteristic (ROC) curve and forest plot describing test accuracy studies of (A) Folstein’s mini-mental state examination (MMSE) at threshold of <25/30; and (B) MMSE at threshold of <27/30.
Figure 3 (Continued). Summary ROC curve and forest plot describing test accuracy studies of (C) Montreal cognitive assessment (MoCA) at threshold of <26/30; and (D) MoCA at threshold of <22/30. The filled circle represents the summary (pooled) test accuracy; unfilled circles represent individual articles. The solid line is the summary ROC curve; the broken line is the 95% confidence interval (CI) around the summary point.
that a test suitable for longer-term poststroke assessment may not be suitable for early assessment and that cognitive testing can (and should) be performed in the acute stroke unit.

Brief assessment tools are attractive for use in busy stroke settings but only if they have acceptable accuracy. In this regard, the limited number of studies looking at brief assessments is unfortunate. Based on data available, screening using clock drawing test or abbreviated mental test may have a role for initial assessment; but they should not be considered a substitute for subsequent multidomain testing.

Limitations of Included Studies
Our review highlights issues in the design and reporting of cognitive test accuracy studies. There was little reporting of missing or indeterminate data; however, we know that substantial numbers of stroke patients are unable to complete multidomain screening tools. The same concern holds for a reference standard based on extensive neuropsychological testing. Limiting test accuracy data to those able to complete testing will bias results, tending to inflate test accuracy. We would encourage use of the intention to diagnose approach, where the traditional 2×2 test accuracy table is expanded with cells representing those unable to complete index test and reference standard.

A related issue is around generalizability of the included subjects. In our quality assessment, we scored several articles as inappropriate exclusion, including the studies that excluded patients with moderate to severe aphasia or inability to consent. We recognize the challenges of cognitive testing in this group, who by definition have at least single domain impairment; however, excluding patients with communication problems or frank confusion from test accuracy research will limit external validity.

Strengths and Limitations of Review
Our review provides a contemporary synthesis of the rapidly evolving field of cognitive screening in stroke. We offer a sensitive search strategy following best practice in conduct and reporting and with multiple embedded internal and external validity checks. The review has limitations; our focused question excluded potentially informative articles, for example where screening tests are compared against each other. We did not assess other important test metrics, such as responsiveness to change or reliability, or other cognitive states, such as (single domain) mild cognitive impairment. These questions would need their own review. We recognize that cognitive assessment in stroke is evolving; 15 (47%) of included articles in our primary review were published since 2010. Novel, stroke cognition assessments are emerging, but no test accuracy data were available at time of review.

Future Research
Our subgroup analysis suggests that test accuracy will vary depending on time since stroke, but could not suggest an optimal time for assessment. Early assessment in the Acute Stroke Unit has practical advantages and could allow for timely intervention; however, few studies assessed patients in the acute period. Given the issues with generalizability and missing data, future studies may wish to describe feasibility and acceptability of testing, as well as classical test accuracy metrics.

Conclusions
All our results must be interpreted with caution as included studies had substantial heterogeneity and potential for bias. However, we can offer some practical advice to clinicians. There is no clearly superior cognitive screening test approach and no evidence that newer or longer screening tests necessarily offer better test accuracy than established screens with shorter administration time. The strategy for cognitive testing in stroke must be informed by the purpose of the test and by other metrics, such as feasibility, acceptability, and opportunity cost. If the purpose of the initial screening is to pick-up all potential cases to allow further assessment, then MoCA may be the preferable test because it offers high sensitivity with shorter administration time than other equally sensitive tests. For MoCA use in stroke, screen positive thresholds may need to be adapted if the aim is to assess for dementia/multidomain impairment, and test accuracy may vary with time since stroke event.

Acknowledgments
We are grateful for expert assistance from Cochrane Dementia and Cognitive Improvement Group to Cochrane authors who shared relevant papers, specifically Dr Ingrid Arevalo Rodriguez, Dr Sarah Cullum/Dr Daniel Davis, and Professor Nadina Lincoln. We are grateful to all study authors who responded to queries or shared data.

Sources of Funding
Project supported by a Chest Heart and Stroke Scotland research award. J. Selvarajah is supported by a National Health Service Research Scotland fellowship, Chief Scientists Office Scotland. D.J. Stott has received support toward cognitive assessment research from Chief Scientists Office and the Orr Bequest. T.J. Quinn received travel support from Royal College of Physicians and Surgeons Glasgow Scholarship and Graham Wilson Travel Fund.

Disclosures
None.

References


Test Accuracy of Cognitive Screening Tests for Diagnosis of Dementia and Multidomain Cognitive Impairment in Stroke
Rosalind Lees, Johann Selvarajah, Candida Fenton, Sarah T. Pendlebury, Peter Langhorne, David J. Stott and Terence J. Quinn

Stroke. published online September 4, 2014;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2014/09/04/STROKEAHA.114.005842

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2014/09/19/STROKEAHA.114.005842.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org/subscriptions/
Supplementary Materials

I Detailed summary of search strategy
II Papers used to assess external validity of search
III Summary of included studies
IV QUADAS-2 based assessment at individual study level
V STARDdem descriptors of reporting
VI Forest plots of test accuracy data
VII Summary ROC exploring effect of covariates (setting and reference standard)
VIII Summary of studies describing brief cognitive tests
Supplemental Methods

I Search strategy

We searched the following databases from inception to January 2014. We applied no language or date restrictions.

ALOIS (Cochrane Dementia and Cognitive Improvement Group); ARIF (University of Birmingham); CINAHL (EBSCOhost); Embase (OvidSP); LILACS (Bireme); Medline (OvidSP); MEDION (Netherlands); Psychinfo (OvidSP) and the DARE, NHS EED, HTA databases (CRD).

We handsearched recent publications (2010 onwards) in key journals including conference proceedings (Age and Ageing; Cerebrovascular Diseases; International Journal of Stroke; Lancet Neurology; Stroke).

We contacted groups with research interest in stroke test accuracy. We utilised “related article” feature in PubMed and examined key studies in the citation databases of Science Citation Index and Scopus.

We supplemented our sensitive search with a purposive search, focussed on four prevalent cognitive screening tools: AMT, MMSE, Montreal Cognitive Assessment (MoCA) and Addenbrookes’ Cognitive Examination Revised (ACE-R).
Search strategy (as used in “Medline”)

Concept a) Stroke
1. Brain Ischemia.ti,ab (exploded)
2. Cerebrovascular*ti,ab (exploded)
3. Stroke.ti,ab (exploded)
4. or/1-3

Concept b) Cognitive disorders and tests
1. 3-stage commands.ti,ab
2. Abbreviated mental test.ti,ab
3. AMT.ti,ab
4. Addenbrooke’s cognitive examination revised.ti,ab
5. ACE-R.ti,ab
6. Arizona battery for communication disorders of dementia.ti,ab
7. Assessment of motor and processing skills.ti,ab
8. AMPS.ti,ab
9. Block tapping.ti,ab
10. Brixton tests.ti,ab
11. California verbal learning test.ti,ab
12. CVLT.ti,ab
13. Cambridge cognitive examination revised.ti,ab
14. CAMCOG.ti,ab
15. Chessington occupational therapy Neurological assessment battery.ti,ab
16. COTNAB.ti,ab
17. Clock drawing.ti,ab
18. Cognitive linguistic quick tester.ti,ab
19. CLQT.ti,ab
20. Cognistat.ti,ab
21. Doors and people test.ti,ab
22. Galveston orientation and amnesia test. ti, ab
23. GOAT. ti, ab
24. Hayling test. ti, ab
25. Intersecting pentagons. ti, ab
26. Loewenstein occupational therapy cognitive assessment. ti, ab
27. LOTCA-G. ti, ab
28. Lothian aphasia stroke cognitive assessment. ti, ab
29. LASKA. ti, ab
30. Mental status questionnaire. ti, ab
31. MSQ. ti, ab
32. Mini mental state examination. ti, ab
33. MMSE. ti, ab
34. Montreal cognitive assessment. ti, ab
35. MoCA. ti, ab
36. Measure of cognitive linguistic ability. ti, ab
37. MCLA. ti, ab
38. OT cognitive screening tool. ti, ab
39. Perceive, recall, Plan and Perform. ti, ab
40. PRPP. ti, ab
41. Picture cards. ti, ab
42. Repeatable battery for the assessment of the neuropsychological status. ti, ab
43. RBANS. ti, ab
44. Rivermead behavioural memory test. ti, ab
45. RBMT. ti, ab
46. Rivermead perceptual assessment battery. ti, ab
47. RPAB. ti, ab
48. Screening Instrument for neuropsychological Impairments in Stroke. ti, ab
49. SINS. ti, ab
50. Short orientation memory and concentration test. ti, ab
51. SOMC.ti,ab
52. Verbal fluency test .ti,ab
53. Wessex head injury matrix.ti,ab
54. WHIM.ti,ab
55. Alzheimer Disease.ti. ab (exploded)
56. Cognition.ti. ab
57. Cognition Disorders.ti,ab (exploded)
58. Dementia.ti,ab (exploded)
59. Memory.ti,ab (exploded)
60. Vascular dementia.ti,ab

61. or/1-54
62. or/55-60
63. 61 or 62

**Concept c) cognitive screening**
1. Mass Screening.ti,ab (exploded)
2. Mental Status Schedule.ti,ab
3. Neuropsychological Tests.ti,ab (exploded)
4. Predictive Value of Tests.ti,ab
5. Psychiatric Status Rating Scales.ti,ab (exploded)
6. Psychological Tests.ti,ab
7. Reproducibility of Results.ti,ab
8. ROC Curve.ti,ab
9. Sensitivity.ti,ab (exploded)
10. Specificity.ti,ab (exploded)
11. Severity of Illness Index.ti,ab (exploded)
12. or/1-11

(Concept b OR concept c) AND concept a
Il Ten papers used for external validity check of search strategy


Supplemental tables

Table III Summary of included studies

a) Limited to “acute” settings
b) Limited to “non-acute” settings
<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>Time since stroke</th>
<th>Includes aphasia</th>
<th>Index test(s)</th>
<th>Index test rater</th>
<th>Diagnostic test(S)</th>
<th>Diagnostic test rater</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bour 2010¹</td>
<td>Neurology ward</td>
<td>Acute</td>
<td>no</td>
<td>MMSE</td>
<td>Psychologist</td>
<td>NPB</td>
<td>Psychologist</td>
</tr>
<tr>
<td>Cumming 2010²</td>
<td>Unspecified</td>
<td>Acute</td>
<td>unspecified</td>
<td>MMSE, MoCA, MMSE</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Dong 2012³</td>
<td>ASU</td>
<td>Hyperacute</td>
<td>no</td>
<td>MoCa, MMSE</td>
<td>Not specified</td>
<td>NPB</td>
<td>Psychologist</td>
</tr>
<tr>
<td>Dong 2010⁴</td>
<td>ASU</td>
<td>Hyperacute</td>
<td>no</td>
<td>MoCa, MMSE</td>
<td>Not specified</td>
<td>Clinical</td>
<td>Not specified</td>
</tr>
<tr>
<td>Godefroy 2011⁵</td>
<td>ASU</td>
<td>Acute</td>
<td>yes</td>
<td>MoCa, MMSE</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Green 2013⁶</td>
<td>ASU</td>
<td>Acute</td>
<td>unspecified</td>
<td>RBANS</td>
<td>Researcher</td>
<td>NPB</td>
<td>Researcher</td>
</tr>
<tr>
<td>Jodzio 2010⁷</td>
<td>Neurology ward</td>
<td>Acute</td>
<td>no</td>
<td>WCST</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Morris 2012⁸</td>
<td>ASU</td>
<td>Acute</td>
<td>yes (mild)</td>
<td>ACE-R, MMSE</td>
<td>Physician</td>
<td>NPB</td>
<td>Psychologist</td>
</tr>
<tr>
<td>Nys 2005⁹</td>
<td>ASU</td>
<td>Acute</td>
<td>yes</td>
<td>MMSE</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Salvadori 2013¹⁰</td>
<td>ASU</td>
<td>Hyperacute*</td>
<td>yes (mild)</td>
<td>MoCA</td>
<td>Researcher</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Wu 2013¹¹</td>
<td>ASU</td>
<td>Unspecified</td>
<td>no</td>
<td>MoCA</td>
<td>Physician</td>
<td>Clinical</td>
<td>Not specified</td>
</tr>
<tr>
<td>Study</td>
<td>Setting</td>
<td>Time since stroke</td>
<td>Includes aphasia</td>
<td>Index test(s)</td>
<td>Index test rater</td>
<td>Diagnostic test(S)</td>
<td>Diagnostic test rater</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Agrell 2000</td>
<td>Rehabilitation</td>
<td>Post acute</td>
<td>no</td>
<td>MMSE</td>
<td>Physician</td>
<td>NPB</td>
<td>Psychologist</td>
</tr>
<tr>
<td>Baum 2008</td>
<td>Community</td>
<td>Post acute</td>
<td>unspecified</td>
<td>EFPT</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Blake 2002</td>
<td>Hospital (other)</td>
<td>Unspecified</td>
<td>unspecified</td>
<td>MMSE, SST, RCPM</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Brookes 2012</td>
<td>Rehabilitation</td>
<td>Post acute</td>
<td>unspecified</td>
<td>BMET, MMSE, CDR</td>
<td>Not specified</td>
<td>Clinical</td>
<td>Not specified</td>
</tr>
<tr>
<td>Cartoni 2007</td>
<td>Rehabilitation</td>
<td>Unspecified</td>
<td>no</td>
<td>MEAMS</td>
<td>OT</td>
<td>NPB</td>
<td>Psychologist</td>
</tr>
<tr>
<td>Cumming 2013</td>
<td>Community</td>
<td>Long term</td>
<td>yes</td>
<td>MMSE, Cog4</td>
<td>Physician Psychiatrist</td>
<td>Clinical</td>
<td>Physician Psychiatrist</td>
</tr>
<tr>
<td>de Koning 2000</td>
<td>Neurology ward</td>
<td>Medium term</td>
<td>no</td>
<td>(R)-CAMCOG</td>
<td>Physician Psychiatrist</td>
<td>Clinical</td>
<td>Adjudication panel</td>
</tr>
<tr>
<td>de Koning 2005</td>
<td>Hospital (other)</td>
<td>Medium term</td>
<td>no</td>
<td>R-CAMCOG</td>
<td>Researcher</td>
<td>Clinical</td>
<td>Adjudication panel</td>
</tr>
<tr>
<td>de Koning 1998</td>
<td>Outpatients</td>
<td>Post acute</td>
<td>yes (mild)</td>
<td>CAMCOG, MMSE</td>
<td>Not specified</td>
<td>Clinical</td>
<td>Adjudication panel</td>
</tr>
<tr>
<td>Desmond 1994</td>
<td>Outpatients</td>
<td>Unspecified</td>
<td>unspecified</td>
<td>TICS</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Grace 1995</td>
<td>Rehabilitation</td>
<td>Unspecified</td>
<td>no</td>
<td>MMSE, 3MS</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Hershey 1987</td>
<td>Outpatients</td>
<td>unspecified</td>
<td>no</td>
<td>CCSE, FAQ</td>
<td>Mixed</td>
<td>Clinical</td>
<td>Physician</td>
</tr>
<tr>
<td>Larson 2005</td>
<td>Rehabilitation</td>
<td>Medium term</td>
<td>yes (mild)</td>
<td>RBANS</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Nokleby 2008</td>
<td>Rehabilitation</td>
<td>Post acute</td>
<td>yes</td>
<td>Cognistat, SINS, CDT</td>
<td>Physician Psychiatrist</td>
<td>NPB</td>
<td>Psychologist</td>
</tr>
<tr>
<td>Pendlebury 2012</td>
<td>Outpatients</td>
<td>Long term</td>
<td>no</td>
<td>MMSE, MoCA, ACE-R</td>
<td>Physician</td>
<td>NPB</td>
<td>Mixed</td>
</tr>
<tr>
<td>Srikanth 2006</td>
<td>Outpatients</td>
<td>Post acute</td>
<td>yes</td>
<td>S-MMSE</td>
<td>Not specified</td>
<td>Clinical</td>
<td>Not specified</td>
</tr>
<tr>
<td>Tang 2005</td>
<td>Outpatients</td>
<td>Post acute</td>
<td>unspecified</td>
<td>MDRS, MMSE</td>
<td>Researcher</td>
<td>Clinical</td>
<td>Psychiatrist</td>
</tr>
<tr>
<td>Wolf 2010</td>
<td>Rehabilitation</td>
<td>Hyperacute</td>
<td>yes</td>
<td>EFPT</td>
<td>Not specified</td>
<td>NPB</td>
<td>Not specified</td>
</tr>
<tr>
<td>Wong 2009</td>
<td>Rehabilitation</td>
<td>Post acute</td>
<td>unspecified</td>
<td>MoCA</td>
<td>Researcher</td>
<td>NPB</td>
<td>Researcher</td>
</tr>
</tbody>
</table>
N/A=not applicable

MCI=Mild Cognitive Impairment; NPB=Neuropsychological battery

BMET= Brief Memory and Executive Test; CDT=Clock drawing test; CSCSE=Cognitive Capacity Screening Examination; EPFT=Executive Performance Function Test; MEAMS=Middlesex Elderly Assessment Mental State; MMSE=Mini-mental State Examination; MoCA=Montreal Cognitive Assessment; PNB=Preliminary Neuropsychological Battery; RBANS=Repeatable Battery for Assessment of Neuropsychological Status; R-CAMCOG=Rotterdam CAMCOG; SINS=Screening Instrument Neuropsychological impairments in Stroke; TICS=Telephone Interview Cognitive Status; WCST= Wisconsin Card Sorting Test
### Table IV

Summary of studies describing brief (less than 5 minutes administration time

<table>
<thead>
<tr>
<th>Study</th>
<th>&quot;n&quot; included</th>
<th>Setting (Timing)</th>
<th>Index tests</th>
<th>Reference standard(s)</th>
<th>Summary results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson-Greene 2009</td>
<td>115</td>
<td>ASU (Hyperacute)</td>
<td>TCT</td>
<td>MMSE, HVLT</td>
<td>TCT correlates with HVLT and MMSE and discriminates cases from controls</td>
</tr>
<tr>
<td>Lees 2013</td>
<td>111</td>
<td>ASU (Hyperacute)</td>
<td>4AT, AMT, CDT, Cog-4</td>
<td>MoCA</td>
<td>4AT performs well at standard MoCA thresholds CDT performs well at lower MoCA thresholds</td>
</tr>
<tr>
<td>Wong 2004</td>
<td>68</td>
<td>Outpatients (Unspecified)</td>
<td>CDT</td>
<td>MMSE, WCST</td>
<td>CDT correlates with MMSE and WCST and discriminates cases from controls</td>
</tr>
</tbody>
</table>

ASU=Acute Stroke unit

4AT=4 A test; AMT=Abbreviated mental test; CDT=Clock drawing test; TCT=Three Cities Test

HVLT=Harvard Verbal Learning Test; MMSE=Mini Mental State Examination; MoCA=Montreal Cognitive Assessment; WCST=Wisconsin Card Sorting Test
Supplemental Figures

Figure V Risk of bias and applicability concerns assessed at study level using the revised Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2).
<table>
<thead>
<tr>
<th>Risk of Bias</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Selection</td>
<td>Index Test</td>
</tr>
<tr>
<td>Agrell 2000</td>
<td>🔴</td>
</tr>
<tr>
<td>Baum 2008</td>
<td>🔴</td>
</tr>
<tr>
<td>Brookes 2012</td>
<td>🔴</td>
</tr>
<tr>
<td>Cumming 2010</td>
<td>🔴</td>
</tr>
<tr>
<td>de Koning 1998</td>
<td>🔴</td>
</tr>
<tr>
<td>de Koning 2000</td>
<td>🔴</td>
</tr>
<tr>
<td>de Koning 2005</td>
<td>🔴</td>
</tr>
<tr>
<td>Desmond 1994</td>
<td>🔴</td>
</tr>
<tr>
<td>Dong 2012</td>
<td>🔴</td>
</tr>
<tr>
<td>Godefroy 2011</td>
<td>🔴</td>
</tr>
<tr>
<td>Grace 1995</td>
<td>🔴</td>
</tr>
<tr>
<td>Green 2013</td>
<td>🔴</td>
</tr>
<tr>
<td>Hershey 1987</td>
<td>🔴</td>
</tr>
<tr>
<td>Hobson 2003</td>
<td>🔴</td>
</tr>
<tr>
<td>Jodzio 2010</td>
<td>🔴</td>
</tr>
<tr>
<td>Larson 2005</td>
<td>🔴</td>
</tr>
<tr>
<td>Morris 2012</td>
<td>🔴</td>
</tr>
<tr>
<td>Nokleby 2008</td>
<td>🔴</td>
</tr>
<tr>
<td>Nys 2005</td>
<td>🔴</td>
</tr>
<tr>
<td>Pendlebury 2012</td>
<td>🔴</td>
</tr>
<tr>
<td>Salvadori 2013</td>
<td>🔴</td>
</tr>
<tr>
<td>Sodring 1998</td>
<td>🔴</td>
</tr>
<tr>
<td>Srikanth 2006</td>
<td>🔴</td>
</tr>
<tr>
<td>Wolf 2010</td>
<td>🔴</td>
</tr>
<tr>
<td>Wong 2009</td>
<td>🔴</td>
</tr>
<tr>
<td>Wu 2013</td>
<td>🔴</td>
</tr>
</tbody>
</table>

- 🔴: High
- ?: Unclear
- 🔴: Low
### Figure VI STARDdem reporting guidance as applied to included studies

| STARIdem item | 1 2 25 | 3 4 5 6 7 8 9 10 11 12 13 | 14 15 16 17 18 19 20 21 22 23 24 |
|---------------|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Agrell 2000   | N Y Y Y N Y Y Y Y Y Y Y Y Y Y | N Y Y Y N N Y N Y Y | N N Y N Y N Y N Y Y | N N Y N Y N Y N Y Y |
| Baum 2008     | Y Y Y Y Y Y N Y Y Y Y N Y Y | Y Y N Y Y N N Y N N Y | Y N N Y N N Y N N Y | Y N N Y N N Y N N Y |
| Blake 2002    | N Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Bour 2010     | N Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Brookes 2012  | N Y Y Y Y N Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Cartoni 2007  | Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Cumming 2010  | N Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Cumming 2013  | N Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| de Koning 2000| N Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| de Koning 2005| N Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| de Koning 1998| N Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Desmond 1994  | Y Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Dong 2012     | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Dong 2010     | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Godefroy 2011 | N Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Grace 1995    | Y Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Green 2013    | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Hershey 1987  | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Hobson 2003   | N Y Y Y Y N Y Y Y Y Y Y N Y Y | N Y N Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Jodzio 2010   | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Larson 2005   | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Morris 2012   | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Nkley 2008    | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Nys 2005      | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Pendlebury 2012| N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Salvadori 2013| N Y Y Y Y N Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Srikant 2006  | N Y Y Y Y N Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Tang 2005     | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Wolf 2010     | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Wong 2009     | N Y Y Y Y N Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
| Wu 2013       | N Y Y Y Y Y Y Y Y Y Y Y Y Y Y | N Y Y Y Y N N Y N N Y | N N Y N Y N Y N Y N | N N Y N Y N Y N Y N |
Figure VII Forest plots of test accuracy data for cognitive screening tools in stroke

Forest plot of ACE-R (threshold 88) for diagnosis of dementia / cognitive impairment (n=2 studies)

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morris 2012</td>
<td>34</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>0.94 [0.81, 0.99]</td>
<td>0.80 [0.69, 0.92]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pendlebury</td>
<td>18</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0.54 [0.39, 0.79]</td>
<td>0.95 [0.73, 0.99]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Forest plot of R-CAMCOG (threshold 25) for dementia / cognitive impairment (n=2 studies)

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>de Koning A</td>
<td>50</td>
<td>23</td>
<td>5</td>
<td>206</td>
<td>0.91 [0.90, 0.92]</td>
<td>0.90 [0.85, 0.94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Koning B</td>
<td>23</td>
<td>5</td>
<td>12</td>
<td>81</td>
<td>0.66 [0.48, 0.81]</td>
<td>0.84 [0.67, 0.91]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure VIII Summary ROC curves exploring effect of co-variates on test accuracy
a) Effect of time since stroke, comparing “acute” testing with “non-acute” testing

The filled circle/diamond represents the summary (pooled) test accuracy; unfilled circles/diamonds represent individual papers. The solid line is the summary ROC curve; the broken line is the 95% confidence interval around the summary point.
b) Effect of reference standard employed, comparing “clinical” diagnosis of dementia against diagnosis made using “neuropsychological battery”

The filled circle/diamond represents the summary (pooled) test accuracy; unfilled circles/diamonds represent individual papers. The solid line is the summary ROC curve; the broken line is the 95% confidence interval around the summary point.
REFERENCES


