Title: Diagnosis in vascular dementia, applying 'Cochrane diagnosis rules' to 'dementia diagnostic tools'
(Editorial on Biesbroek et al)

Running title: Dementia test accuracy

Authors: Terence J Quinn, Jenny McCleery
1. Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK
2. Cochrane Dementia and Cognitive Improvement Group, Oxford, UK

Address for correspondence:
Institute of Cardiovascular and Medical Sciences
University of Glasgow
Room 2.55, New Lister Building, Glasgow Royal Infirmary Campus
Glasgow G4 0SF
Tel: +44 (0)141 201 8510
Email: terry.quinn@glasgow.ac.uk

Keywords: biomarkers; dementia; sensitivity/specificity; test accuracy; vascular cognitive impairment

Diagnosis in vascular dementia, applying 'Cochrane diagnosis rules' to 'dementia diagnostic tools'

"Doctor, will I get dementia?" is a frequent question in the memory clinic. Unfortunately, current assessment tools are often unable to give the patient a satisfactory answer. In clinical practice and in research, making an early diagnosis of dementia, especially vascular dementia, is challenging.[1] A particular issue is that the traditional 'gold standard' diagnostic assessment is neuropathology, which is only really possible in post-mortem samples.[2] Advances in technology are helping us overcome this problem and increasingly sophisticated neuroimaging can give us impressive visualisations of cerebral structure and function in-vivo.

In this issue of Clinical Science, Biesbroek and colleagues describe recent work on Magnetic Resonance Imaging (MRI) based cerebral lesion location and its association with cognitive decline.[3] The authors conclude that diagnostic neuroimaging in dementia should shift from whole brain evaluation to focussed quantitative analysis of strategic brain areas. The review provides a fascinating insight into the evolution of the lesion location concept from single case-study through to lesion-symptom maps based on quantitative analysis of images from large observational cohorts. The authors speculate on the clinical applications that might follow from lesion-symptom mapping in the next few years. Their most immediate ambition for the technique is "to improve the diagnostic work-up of memory clinic patients with SVD (small vessel disease)". Looking further ahead, they hope that the techniques they describe may have prognostic utility, predicting decline in individual cognitive domains and treatment response by enabling the tailoring of treatments to the neurochemistry of affected tracts and regions.

In the field of vascular dementia, or indeed dementia generally, where standard assessment is an imperfect guide to diagnosis and prognosis, it is tempting to embrace new tools in the hope that any development must be better than current practice. However, there is a risk that enthusiasm for new technologies can outpace the data.[4] Advances in technology will almost certainly transform the clinic cognitive work-up, but new assessment tools should be scutinised with the same rigour we would expect to be applied to a new treatment. In the Cochrane Dementia and Cognitive Improvement Group (CDCIG), we strive to collate, synthesise and critically appraise the growing literature around new dementia biomarkers.[5] Since the aim of biomarker-based tests is often to identify patients in earlier stages of disease than has previously been possible, we have, in this context, considered early diagnosis to be inseparable from prognosis. When the clinical question we seek to answer is 'which patients will develop dementia?', then the diagnostic gold standard is the longitudinal outcome.[6] Informed by our systematic reviews, we have produced best practice guidance for the conduct and reporting of studies evaluating potential diagnostic tests for dementia .[7,8]

For randomised control trials, the standardised approach to reporting outlined in the CONSORT (Consolidated Standards of Reporting Trials) statement and checklist has helped raise awareness of methodological and reporting standards.[9] For many scientific journals a completed CONSORT checklist is now mandatory when submitting a paper describing a trial. A similar set of reporting guidelines for diagnostic test accuracy studies is available in the STARD (Standards for Reporting Diagnostic Accuracy) guideline.[10] While STARD is a useful tool, applying these recommendations to studies in dementia can prove problematic.[11] To make a resource more suited to those working in the dementia field, a STARD extension specific to dementia test accuracy studies is now available

(STARDdem).[7] Although primarily a reporting guideline, STARDdem materials can also highlight critical issues at the earlier stages of study design and execution. Some issues which may be of particular importance to future studies of lesion localisation in SVD are the representativeness of the sample, selection of test thresholds, patient flow and analysis methods.

Representativeness of study samples: There is sometimes an apparent disconnect between the participants included in dementia biomarker studies and the population of interest. For example, although early onset monogenic dementias such as CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leucoencephalopathy) offer an excellent opportunity to study a pure dementia phenotype, these rare syndromes are not representative of the older adults with memory problems that we see in practice.[12] For an assessment of a biomarker to have external validity the sampling frame should approximate to the population in which the test will be used. Capturing the 'messy reality' of the patients seen in secondary care memory services, who often have substantial comorbidity and frailty, in a controlled research study is challenging, but is a vital step before changes to practice can be recommended.[13] In this context, Biesbroek and colleagues' point out that assessment of regional lesion volumes should be complemented by biomarker tests assessing other possible causes of brain injury and neurodegeneration. Any single biomarker test may be markedly less informative in older patients with multiple brain pathologies than in younger patients with 'purer' neuropathologies.

Also important for external validity is a move away from the case-control methods used in many of the studies in the Biesbroek review. The case control approach is acceptable for initial validation work, but has many inherent limitations. It typically manipulates the occurrence of the condition of interest (in this case vascular dementia) to enrich the sample with more 'cases'. As the resulting sample does not have the natural population frequency of disease, no comment can be made on metrics that are dependent on disease prevalence such as predictive values.[13] Further, the case-control methodology, by tending to favour phenotypic extremes such as healthy controls with no comorbidity versus patients with frank dementia, artificially inflates measures of test accuracy and does not tell us about the patients of greatest interest – those in whom the clinical diagnosis is uncertain.[14]

A final step to consider around external validity is the need for replication of results in independent cohorts. It is notable that many of the studies included in the Biesbroek review come from the same academic groups. This is not a criticism, as early phase work using cutting edge technology will necessarily be restricted to specialist centres. However, before recommendations can be made around general usage of a test, the positive results seen in early adopting centres need to be confirmed in work from other international sites.

Test thresholds and patient flow: STARD (and STARDdem)[7,10] guidance emphasises the need to report potential sources of bias which can challenge the internal validity of diagnostic test accuracy studies. Some of the issues, such as recruitment and blinding, are well known. However, others may be less intuitive, including the importance of describing test thresholds and patient flow.

For most tests, the output is a series of data that are then assessed to give a binary output of disease/no disease status. To make this classification, a threshold is set beyond which the subject is

said to be ‘test positive’. In Biesbrock's clinical example, this would be the lesion volume in the relevant strategic tract which is considered sufficient to count as an explanation for the patient's symptoms. If test thresholds are not pre-specified in a diagnostic study, then it is tempting to use the threshold that gives the most impressive results and thus artificially increases test accuracy. If various centres all use differing thresholds then comparison becomes impossible.

For any test, not all subjects will complete the test and give usable test data. For novel neuroimaging sequences with potentially long image acquisition times , test non-completion may be a particular concern. Test accuracy reporting should allow for an ‘intention to diagnose’ approach, so that it is immediately apparent how many subjects did not complete the protocol or gave indeterminate/ unusable data.[15] Simply excluding these test ‘drop-outs’ from the analysis will, again, artificially improve apparent test accuracy.[16]

Analysis methods - moving beyond correlations: There are many potential approaches to the quantitative description of how a test performs. A common analysis is to describe correlation between test data and an outcome of interest. Many of the papers included in the Biesbroek review describe correlation of lesion pathology with scores on a neuropsychological battery. Correlation-based analyses are suitable for early phase work with a new diagnostic test. However, for understanding clinical utility, correlation is a fairly blunt instrument and other statistical approaches are preferred.

There is no perfect method for describing test properties. While the paired values of sensitivity and specificity are the most commonly used measures of test accuracy, metrics such as predictive value, which quantify the probability of presence/absence of the target condition given a particular test result, may have greater clinical utility for assessment of an individual patient.[17] The reporting of a 2x2 table, cross-classifying test results with disease status, allows for calculation of sensitivity/specificity, predictive values and many other test metrics and is recommended for all papers describing the properties of diagnostics tests.

Test accuracy metrics should not be interpreted in isolation. Increasingly we are recognising the importance of describing the full test-to-treatment pathway.[18] Performance of a test can be associated with various direct and indirect effects. Test accuracy is not synonymous with clinical effectiveness and an accurate test does not necessarily result in improved patient outcomes. A particular concern when testing for a neurodegenerative condition such as dementia is the potential consequence of test error. If a novel imaging test gives a patient with early or subclinical dementia a false negative result, this is unfortunate, but probably of relatively little consequence. The natural history of preclinical dementia states is highly variable, at present we have no proven preventative treatment and the true diagnosis will emerge when the subject becomes symptomatic. However, a misdiagnosis of incipient dementia in a healthy individual will likely have substantial negative effects on their psychological health and lead to potentially inappropriate resource use through onward referrals, further testing and follow-up. Many of the cerebrospinal fluid biomarkers tests for dementia, for example, have a false positive rate that, although modest, may still be too high for routine use in clinical practice.[19]

Moving beyond accuracy, aspects of test evaluation including test feasibility and acceptability are important to understand before advocating routine use of a test. To date these metrics have been rarely described for dementia biomarkers. Where a new test is being considered in favour of an existing approach it can be useful to describe the incremental benefit over standard practice. For example, in a recent study looking at imaging biomarkers, the authors found reasonable test accuracy of the biomarkers, but when considered in the context of standard memory testing there was little additional value of these sophisticated tests (calculated using a net re-classification index).[20]

The next generation of neuroimaging biomarker studies: These comments are not intended as criticisms of the existing research on imaging biomarkers, but should draw attention to the challenges facing future studies. For lesion location work, Biesbroek's review demonstrates that proof of concept is now established. The next steps towards clinical utility should build on the considerable experience accumulated in studies of biomarkers for the diagnosis of Alzheimer's disease. Proposed revisions of diagnostic criteria for dementia place increasing emphasis on biomarkers.[21] We argue here that enthusiasm for the new must be tempered with a keen critical appraisal in order to maximise the benefits of new technologies while avoiding premature adoption of diagnostic strategies that are no better than standard approaches or even have potential unintended harms.

Dr Quinn and McCleery are the joint coordinating editors of the Cochrane Dementia and Cognitive Improvement Group (http://dementia.cochrane.org/) and have published guidelines and best practice statements on dementia test accuracy studies.

References

1. Scheltens P, Blennow K, Bretele MMB et al. Alzheimer's disease. Lancet. 2016; 388:505-517.

2. Beach TG, Onsell SE, Philips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer Disease at National Institute on Ageing Alzheimer's Disease Centres 2005-2010. J Neuropath Exp Neurol. 2012; 71:266-273.

3. Biesbroek JM, Weaver NA, Biessels GJ. Lesion location and cognitive impact of small vessel disease. Clinical Science in press.

4. Dyer SM, Flicker L, Laver K, Whitehead C, Cumming R. The clinical value of fluid biomarkers for dementia diagnosis. Lancet Neurology. 2016;15:1204.

5. Quinn TJ, McCleery J, Noel-Storr AH, Marcus S, Flicker L. Cochrane Dementia Group Turns 21 – older and (slightly) wiser. JAMDA. 2017;18:96-98.

6. Harrison JK, Stott DJ, McShane R, Noel-Storr AH, Swann-Price RS, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early diagnosis of dementia across a variety of healthcare settings. Cochrane Database Syst Rev. 2016;(11):CD011333.

7. Noel-Storr AH, McCleery JM, Richard E, et al. Reporting standards for studies of diagnostic test accuracy in dementia: the STARDdem initiative. Neurology. 2014;83:364–73.

8. Davis DH, Creavin ST, Noel-Storr A, 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 Syst Rev. 2013;(3):CD010460.

9. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust. 2006;185:263–7.

10. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.

11. Noel-Storr AH, Flicker L, Ritchie CW, et al. Systematic review of the body of evidence for use of biomarkers in the diagnosis of dementia. Alzheimers Dement. 2013;9:e96–105.

12. Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology. 2007;69:2197–204.

13. Ritchie CW, Terrera GM, Quinn TJ. Dementia trials and dementia tribulations. Methodological and analytical challenges in dementia research. Alzheimers Res Ther. 2015;7:31.

13. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282(11):1061-1066.

14. Reitsma J, Rutjes A, WP, Vlassov V, Leeflang M, Deeks J. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration; 2009.

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

16. Lees RA, Hendry K, Broomfi eld N, Stott D, Larner AJ, Quinn TJ. Cognitive assessment in stroke: feasibility and test properties using differing approaches to scoring of incomplete items. Int J Geriatr Psychiatry. 2016. doi: 10.1002/gps.4568.[Epub ahead of print].

17.Moons KG, Harrell FE. Sensitivity and specificity should be de-emphasised in diagnostic accuracy studies. Acad Radiol. 2003;10:670–2.

18. Gopalakrishna G, Langendam MW, Scholten RJPM, Bossuyt PMM, Leeflang MMG. Defining the clinical pathway in Cochrane diagnostic test accuracy reviews. BMC Med Res Methodology. 2016. 16:153.

19. Ritchie CW, Smailagic N, Noel-Storr AH et al Plasma and cerebrospinal fluid amyloid beta for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev. 2014:CD008782.

20. Richard E, Schmand BA, Eikelenboom P, Van Gool WA. The Alzheimer's Disease neuroimaging initiative MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: a diagnostic accuracy study. BMJ Open. 2013;3:e002541.

21. Dubois B, Feldman HH, Jacova C, et al. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol. 2010;9:1118–27.