

Original Research Article

Comparing Diagnostic Accuracy of Cognitive Screening Instruments: A Weighted Comparison Approach

A.J. Larner

Cognitive Function Clinic, Walton Centre for Neurology and Neurosurgery, Liverpool, UK

Key Words

Diagnostic accuracy · Weighted comparison · Mini-Mental State Examination · Addenbrooke's Cognitive Examination-Revised · Montreal Cognitive Assessment · Test Your Memory test · Mini-Mental Parkinson

Abstract

Background/Aims: There are many cognitive screening instruments available to clinicians when assessing patients' cognitive function, but the best way to compare the diagnostic utility of these tests is uncertain. One method is to undertake a weighted comparison which takes into account the difference in sensitivity and specificity of two tests, the relative clinical misclassification costs of true- and false-positive diagnosis, and also disease prevalence. **Methods:** Data were examined from four pragmatic diagnostic accuracy studies from one clinic which compared the Mini-Mental State Examination (MMSE) with the Addenbrooke's Cognitive Examination-Revised (ACE-R), the Montreal Cognitive Assessment (MoCA), the Test Your Memory (TYM) test, and the Mini-Mental Parkinson (MMP), respectively. **Results:** Weighted comparison calculations suggested a net benefit for ACE-R, MoCA, and MMP compared to MMSE, but a net loss for TYM test compared to MMSE. **Conclusion:** Routine incorporation of weighted comparison or other similar net benefit measures into diagnostic accuracy studies merits consideration to better inform clinicians of the relative value of cognitive screening instruments.

Copyright © 2013 S. Karger AG, Basel

Introduction

A large number of cognitive screening instruments (CSI) is available for the assessment of patient complaints of poor memory or cognitive impairment [1–3]. Although criteria for the optimal CSI have been suggested [4], in practice many different approaches to comparing tests may be undertaken, essentially balancing test speed against accuracy.

A.J. Larner
Cognitive Function Clinic, Walton Centre for Neurology and Neurosurgery
Lower Lane, Fazakerley
Liverpool L9 7LJ (UK)
E-Mail a.larner@thewaltoncentre.nhs.uk

Table 1. Study demographics

Study	Setting	n	Dementia prevalence, %	M:F n (% male)	Age range, years	Ref.
ACE-R vs. MMSE	cognitive function clinic	243	35	135:108 (56)	24–85 (mean 59.8±10.9)	10
MoCA vs. MMSE	cognitive function clinic	150	24 (MCI 19)	93:57 (62)	20–87 (median 61)	12
TYM vs. MMSE	cognitive function clinic and old age psychiatry memory clinic	224	35	130:94 (58)	20–90 (mean 63.3±12.6)	14
MMP vs. MMSE	cognitive function clinic	201	23	115:56 (57)	20–86 (median 62)	16

MCI = Mild cognitive impairment.

When assessing the diagnostic utility of CSI, a number of summary measures is available, which may help to guide clinicians in selecting the test most appropriate for the purpose. In addition to test sensitivity and specificity, diagnostic utility may be expressed in terms of predictive values, likelihood ratios, clinical utility index, agreement between tests (kappa statistic), and the area under the receiver-operating characteristic curve (AUC). All these measures have potential shortcomings.

AUC is commonly used as an overall measure of diagnostic test accuracy, but the shortcomings of this measure have been emphasized [5], specifically the fact that it combines test accuracy over a range of thresholds which may be both clinically relevant and clinically nonsensical. It has been argued that the most relevant and applicable presentation of diagnostic accuracy test results should include interpretation in terms of patients, clinically relevant values for test thresholds, disease prevalence, and clinically relevant relative gains and losses [5]. One such index is the weighted comparison (WC) measure, described by Moons et al. [6], which gives weighting to the difference in sensitivity and specificity of two tests and takes into account the relative clinical misclassification costs of true-positive (TP) and false-positive (FP) diagnosis and also disease prevalence.

The aim of this study was to reinterpret data from a number of pragmatic prospective diagnostic accuracy studies performed in this clinic [7] in terms of the WC measure of Moons et al. [6] in order to compare a number of CSI, specifically the Mini-Mental State Examination (MMSE) [8] with the Addenbrooke's Cognitive Examination-Revised (ACE-R) [9, 10], the Montreal Cognitive Assessment (MoCA) [11, 12], the Test Your Memory (TYM) test [13, 14], and the Mini-Mental Parkinson (MMP) [15, 16].

Materials and Methods

Data from four previous pragmatic diagnostic accuracy studies undertaken in this clinic, which compared the MMSE with the ACE-R [10], MoCA [12], TYM test [14], and MMP [16], were reanalyzed. Study details (setting, sample size, dementia prevalence, sex ratio, and age range) are given in table 1. Test cutoffs were determined empirically by examining sensitivity and specificity at all cutoff values with the optimal cutoff being defined by maximal test accuracy for diagnosis. In each of these studies, criterion diagnosis was made by the judgment of an experienced clinician based on diagnostic criteria.

Data on the prevalence of dementia [10, 14, 16] or of cognitive impairment (dementia and mild cognitive impairment) [12] were obtained from each study along with the change in test sensitivity and specificity, and these figures were applied to the WC equation [5, 6]:

Table 2. ACE-R vs. MMSE (data adapted from Larner [10])

	ACE-R	MMSE
Sensitivity	0.87	0.70
Specificity	0.91	0.89
AUC	0.94	0.91
PPV	0.83	0.77
NPV	0.93	0.85
Dementia prevalence	0.35	
Δ sensitivity	0.17	
Δ specificity	0.02	
WC	0.17 = net benefit	

PPV = Positive predictive value; NPV = negative predictive value.

$$WC = \Delta \text{ sensitivity} + [(1 - \pi/\pi) \times \text{relative cost (FP/TP)} \times \Delta \text{ specificity}],$$

where π = prevalence.

The relative misclassification cost (FP/TP) is a parameter which seeks to define how many FPs a TP is worth. Clearly, such a ‘cost’ is very difficult to estimate. In the context of diagnostic accuracy studies for CSI, it may be argued that high test sensitivity in order to identify all TPs, with the accompanying risk of FPs (e.g. emotional consequences for a patient because of incorrect diagnosis or inappropriate treatment), is more acceptable than tests with low sensitivity but high specificity which risk false-negative diagnoses (i.e. missing TPs). This argument is of course moot in the current absence of disease-modifying therapies. For this study, FP/TP = 0.1 was therefore arbitrarily set, following previous authors [5], reflecting the desire for high test sensitivity. The WC equation does not take into account false-negative diagnoses, which have their own potential cost.

Positive WC values were taken to indicate a net test benefit, negative values a net loss [5, 6]. To aid interpretation, another parameter may be calculated using WC, namely the equivalent increase in TP patients per 1,000, using the equation [5]:

$$WC \times \text{prevalence} \times 1,000.$$

Again, positive values were taken to indicate a net test benefit, negative values a net loss.

Results

The figures for sensitivity, specificity, prevalence of dementia (ACE-R, TYM, and MMP studies) or cognitive impairment (MoCA study), Δ sensitivity, Δ specificity, and the calculated WC are given for each of the four tests versus MMSE in tables 2–5, respectively. AUC and positive and negative predictive values from each study are included for comparative purposes.

The WC calculations suggested a net benefit for ACE-R, MoCA, and MMP versus MMSE, but a net loss for the TYM test versus MMSE. All WC evaluations were in the same direction as the values for AUC (i.e. favoured ACE-R, MoCA, and MMP vs. MMSE, favoured MMSE vs. TYM test).

The equivalent increase in TP dementia patients identified per 1,000 tested was 61 for ACE-R, –26 for TYM test, and 13 for MMP. The equivalent increase in TP cognitively impaired patients identified per 1,000 tested was 121 for MoCA.

Table 3. MoCA vs. MMSE (data adapted from Larner [12])

	MoCA	MMSE
Sensitivity	0.97	0.65
Specificity	0.60	0.89
AUC	0.91	0.83
PPV	0.65	0.82
NPV	0.96	0.78
Cognitive impairment prevalence	0.43	
Δ sensitivity	0.32	
Δ specificity	-0.29	
WC	0.28 = net benefit	

PPV = Positive predictive value; NPV = negative predictive value.

Table 4. TYM vs. MMSE (data adapted from Hancock and Larner [14])

	TYM	MMSE
Sensitivity	0.73	0.79
Specificity	0.88	0.95
AUC	0.89	0.94
PPV	0.77	0.89
NPV	0.86	0.90
Dementia prevalence	0.35	
Δ sensitivity	-0.06	
Δ specificity	-0.07	
WC	-0.07 = net loss	

PPV = Positive predictive value; NPV = negative predictive value.

Discussion

WC measures may have advantages over more traditional parameters used in the assessment of test utility in diagnostic accuracy studies [6], particularly the AUC. Hence it has been suggested that such measures be incorporated into diagnostic accuracy studies [5]. Net benefit methods to measure test diagnostic performance other than the WC developed by Moons et al. [6] have been described [5].

In this study, data from four previous diagnostic accuracy studies of CSI were reanalyzed to calculate WC. These were pragmatic, observational studies involving unselected patient groups with cognitive complaints of unknown aetiology, rather than experimental studies involving patient groups selected by known diagnostic category, and hence the results should be broadly generalizable since they reflect the idiom of clinical practice [17]. The setting and sample characteristics were broadly equivalent for each of these four studies (table 1).

Overall the calculations suggest a net benefit for ACE-R and, to a lesser extent, MMP for the identification of dementia versus MMSE, and for MoCA for the identification of cognitively impaired patients, with a net loss for the TYM test versus MMSE for the identification of dementia. The equivalent increase for MoCA suggested that fewer than 10 patients needed to be evaluated with this test for 1 additional TP cognitively impaired patient to be identified compared to using the MMSE, concordant with the high sensitivity of this test [11]. Such

Table 5. MMP vs. MMSE (data adapted from Larner [16])

	MMP	MMSE
Sensitivity	0.51	0.45
Specificity	0.97	0.98
AUC	0.89	0.87
PPV	0.83	0.88
NPV	0.87	0.85
Dementia prevalence	0.23	
Δ sensitivity	0.06	
Δ specificity	-0.01	
WC	0.06 = net benefit	

PPV = Positive predictive value; NPV = negative predictive value.

figures may be easier for clinicians to interpret compared to AUC. All WC evaluations were in the same direction as the values for AUC (i.e. favoured ACE-R, MoCA, and MMP vs. MMSE, favoured MMSE vs. TYM test).

Of course, WC values would be different if the case mix seen in these clinical studies had a different disease prevalence, and if a different relative misclassification cost was selected. WC values would fall with higher disease prevalence in the clinic samples. However, empirically a fall in the frequency of patients with dementia and cognitive impairment and an increase in individuals with subjective memory impairment have been observed over time in these clinics [7], perhaps related to governmental directives on dementia issued in the United Kingdom [18]. The setting of the relative misclassification cost (FP/TP) was arbitrary but stringent (10 TPs for 1 FP or 1 TP judged to be worth 0.1 FP). If one accepted more FPs and/or fewer TPs (i.e. a less sensitive, more specific test) the ratio would rise and the WC value would be higher. As previously noted, the WC equation does not take into account false-negative diagnoses, which have their own potential cost.

It might be argued that the sample sizes in these studies (range of $n = 150$ – 243 ; table 1) may mean that they were underpowered. Sample size calculations were not performed, as is sometimes recommended for diagnostic accuracy studies [19], although a pragmatic approach to sample size estimates has suggested that normative ranges for sample sizes may be calculated for common research designs, with anything in the range of 25–400 being acceptable [20].

As shown in this study, application of the WC measure is straightforward, as well as theoretically attractive, for test interpretation [5, 6]. No previous analyses of the diagnostic utility of CSI using this WC method have been identified. The study suggests that there is a case for the routine incorporation of WC or other similar net benefit methods to measure diagnostic test performance into diagnostic accuracy studies.

Disclosure Statement

The author declares no conflicts of interest.

References

- 1 Burns A, Lawlor B, Craig S: Assessment Scales in Old Age Psychiatry, ed 2. London, Martin Dunitz, 2004, pp 33–103.
- 2 Tate RL: A Compendium of Tests, Scales, and Questionnaires. The Practitioner's Guide to Measuring Outcomes after Acquired Brain Impairment. Hove, Psychology Press, 2010, pp 91–270.
- 3 Larner AJ (ed): Cognitive Screening Instruments. A Practical Approach. London, Springer, 2013.
- 4 Malloy PF, Cummings JL, Coffey CE, Duffy J, Fink M, Lauterbach EC, Lovell M, Royall D, Salloway S: Cognitive screening instruments in neuropsychiatry: a report of the Committee on Research of the American Neuropsychiatric Association. *J Neuropsychiatry Clin Neurosci* 1997;9:189–197.
- 5 Mallett S, Halligan S, Thompson M, Collins GS, Altman DG: Interpreting diagnostic accuracy studies for patient care. *BMJ* 2012;344:e3999.
- 6 Moons KG, Stijnen T, Michel BC, Büller HR, Van Es GA, Grobbee DE, Habbema DF: Application of treatment thresholds to diagnostic-test evaluation: an alternative to the comparison of areas under receiver operating characteristic curves. *Med Decis Making* 1997;17:447–454.
- 7 Larner AJ: Dementia in Clinical Practice: A Neurological Perspective. Studies in the Dementia Clinic. London, Springer, 2012.
- 8 Folstein MF, Folstein SE, McHugh PR: 'Mini-Mental State'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–198.
- 9 Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR: The Addenbrooke's Cognitive Examination Revised: a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* 2006;21:1078–1085.
- 10 Larner AJ: Addenbrooke's Cognitive Examination-Revised (ACE-R): pragmatic study of cross-sectional use for assessment of cognitive complaints of unknown aetiology. *Int J Geriatr Psychiatry*, in press.
- 11 Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H: The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53:695–699.
- 12 Larner AJ: Screening utility of the Montreal Cognitive Assessment (MoCA): in place of – or as well as – the MMSE? *Int Psychogeriatr* 2012;24:391–396.
- 13 Brown J, Pengas G, Dawson K, Brown LA, Clatworthy P: Self administered cognitive screening test (TYM) for detection of Alzheimer's disease: cross sectional study. *BMJ* 2009;338:b2030.
- 14 Hancock P, Larner AJ: Test Your Memory (TYM) test: diagnostic utility in a memory clinic population. *Int J Geriatr Psychiatry* 2011;26:976–980.
- 15 Mahieux F, Michelet D, Manificier M-J, Boller F, Fermanian J, Guillard A: Mini-Mental Parkinson: first validation study of a new bedside test constructed for Parkinson's disease. *Behav Neurol* 1995;8:15–22.
- 16 Larner AJ: Mini-Mental Parkinson (MMP) as a dementia screening test: comparison with the Mini-Mental State Examination (MMSE). *Curr Aging Sci* 2012;5:136–139.
- 17 Larner AJ: Pragmatic diagnostic accuracy studies. www.bmj.com/contents/345/bmj.e3999/rr/599970 (accessed October 11, 2012).
- 18 Larner AJ: Impact of the National Dementia Strategy in a neurology-led memory clinic. *Clin Med* 2010;10:526.
- 19 Bachmann LM, Puhan MA, ter Riet G, Bossuyt PM: Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ* 2006;332:1127–1129.
- 20 Norman G, Monteiro S, Salama S: Sample size calculations: should the emperor's clothes be off the peg or made to measure? *BMJ* 2012;345:e5728.