



Comparing, Combining and Converting Screening Instruments

6

Contents

6.1	Comparing Cognitive Screening Instruments	174
6.1.1	Weighted Comparison (WC) and Equivalent Increase (EI)	175
6.1.2	Q* Index	180
6.1.3	Comparing Test Speed Versus Test Accuracy	182
6.1.4	Meta-Analysis: ACE and ACE-R	184
6.1.5	Effect Size (Cohen's d)	185
6.1.6	Correlation	188
6.1.7	Cohen's Kappa Statistic: Test of Agreement	189
6.1.8	Bland-Altman Limits of Agreement	189
6.2	Combining Screening Instruments	191
6.2.1	Combining Cognitive Screening Instruments:	
	MMSE and MoCA	192
6.2.2	Combining Informant and Cognitive Screening Instruments	193
6.2.3	Combining Functional and Cognitive Screening Instruments:	
	IADL Scale and ACE-R; Free-Cog	198
6.3	Converting Cognitive Screening Instrument Scores	199
6.3.1	Linear Regression Equations	199
6.4	Summary and Recommendations	200
	References	201

Abstract

This chapter examines various methods of comparing, combining and converting various screening instruments. Methods of comparing cognitive scales include not only the standard measures of discrimination considered in previous chapters but also weighted comparison and effect size, as well as meta-analysis. The diagnostic utility of combinations of scales, both cognitive and non-cognitive, is also considered, as well as ways to make approximate conversions between the scores of different screening tests.

Keywords

Dementia · Diagnosis · Combinations · Cohen's d · Effect size · Equivalent increase · Limits of agreement · Linear regression equation · Meta-analysis · Weighted comparison

6.1 Comparing Cognitive Screening Instruments

As is evident from the previous two chapters, there are many screening instruments, focusing on either cognitive or non-cognitive domains of function, which may be used in the assessment of patients presenting with cognitive complaints (interpreted in the context of the patient history and examination; Chap. 3). How does one decide which of these instruments should be used, which is optimal? The role of clinician preference should not be underestimated in this choice, but ideally it should be based on some rigorous method of comparison between tests.

One strategy, adopted in previous editions of this book (Larner 2012a:50–2, 97–8; 2014a:131–3, 189–91), is to construct “league tables” for various diagnostic metrics e.g. likelihood ratios, diagnostic odds ratios, clinical utility indexes, and area under the receiver operating characteristic curve (AUC ROC) (see also Larner 2015a, Chap. 4). As was then pointed out, such “league table” comparisons relate to historical and usually non-overlapping patient cohorts, so direct comparisons between instruments cannot be made. Nonetheless, such “league tables” may give some clues as to the relative merits of the tests used in pragmatic studies.

Ideally however, comparison requires head-to-head studies where two (or more) instruments are administered (in random order), and blinded to the result of the other(s), to the cohort of patients undergoing assessment. This is potentially a time-consuming strategy, and fatiguing for patients, although it has been used in some studies performed in the Cognitive Function Clinic (CFC) at the Walton Centre for Neurology and Neurosurgery (WCNN) in Liverpool to compare cognitive screening instruments (CSIs), e.g. MMSE vs. ACE (Larner 2005), MMSE vs. ACE-R (Larner 2009, 2013a), MMSE vs. MoCA (Larner 2012b), MMSE vs. MMP (Larner 2012c), MACE vs. MMSE (Larner 2015a, b), and MACE vs. MoCA (Larner 2017a). Ideally tests should be administered sequentially in counter-balanced order to avoid bias, although this is not possible in some circumstances (e.g. because MMSE is incorporated into both ACE and ACE-R; Sects. 4.1.5.1 and 4.1.5.3).

Other methods of comparison may be based on:

- measures of discrimination (Sect. 2.3.2), as documented for individual screening instruments (Chaps. 4 and 5), such as weighted comparison or Q* index;
- measures based on the reference standard diagnosis, such as effect size (Cohen's *d*);
- or measures of association (non-diagnostic), such as correlation, test of agreement (Cohen's kappa statistic), or Bland-Altman limits of agreement (Sect. 2.3.3).

6.1.1 Weighted Comparison (WC) and Equivalent Increase (EI)

The shortcomings of AUC ROC as an overall measure of diagnostic test accuracy have been emphasized (Mallett et al. 2012), specifically the fact that this unitary metric 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 (Mallett et al. 2012).

One such index is the weighted comparison (WC) measure described by Moons et al. (1997) which gives weighting to the difference in sensitivity and specificity of two tests and takes into account the relative clinical misclassification costs of false positive diagnosis and also disease prevalence. This may be expressed by the equation:

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

where π = prevalence; FP = false positives; and TP = true positives.

The relative misclassification cost (FP/TP) is a parameter which seeks to define how many false positives a true positive is worth. Clearly, such a "cost" is very difficult to estimate. In the context of diagnostic accuracy studies for CSIs, it may be argued that high test sensitivity to identify all true positives, with the accompanying risk of false positives (e.g. emotional consequences for a patient of an incorrect diagnosis, and/or inappropriate treatment), is more acceptable than tests with low sensitivity but high specificity which risk false negative diagnoses (i.e. missing true positives, and possibly the opportunity to initiate symptomatic or disease-modifying treatment). This argument is of course moot in the current absence of disease modifying therapies for most causes of dementia or MCI. For studies in CFC, FP/TP was arbitrarily set at 0.1, following previous authors (Mallett et al. 2012), reflecting the desire for high test sensitivity.

Of note, the WC equation used here (Moons et al. 1997) does not take into account false negative diagnoses, which of course have their own potential cost. However, another index, addressing whether screening tests are "costworthy", also incorporates the benefit (advantage) of TP test for an identified individual and the cost (harm) of FP test for a wrongly identified individual but without reference to false negatives (Ashford 2008).

To aid interpretation, another parameter may be calculated using WC, namely the equivalent increase (EI) in TP patients per 1000, using the equation:

$$EI = WC \times \text{prevalence} \times 1000$$

As this is a measure of patient numbers, results are rounded to integer values.

Weighted comparison and calculation of equivalent increase has been undertaken for a number of the CSIs examined in CFC. These have compared patient performance-related CSIs: MMSE with ACE-R, MoCA, TYM, MMP (Larner 2013b), 6CIT (Abdel-Aziz and Larner 2015) and MACE (Larner 2015a), as well as TYM against ACE-R, and MoCA against MACE (Larner 2016a, 2017a). Comparison of performance-related CSIs and informant scales has also been examined: AD8 with MMSE and 6CIT (Larner 2015c). Most comparisons have been for the diagnosis of dementia, but some also for the diagnosis of MCI. The figures for sensitivity, specificity, AUC ROC, prevalence of dementia/MCI/cognitive impairment were extracted from each study, and Δ sensitivity, Δ specificity, and WC and EI were then calculated (Tables 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8 and 6.9).

The dataset from a patient cohort seen in an old age psychiatry memory clinic (Hancock and Larner 2015; Sect. 5.2.4) permitted a further weighted comparison of MMSE and ACE-R in an independent cohort ($n = 181$) to be undertaken, with results akin to that from the CFC study (Larner and Hancock 2014; compare Tables 6.1 and 6.10).

Table 6.1 Weighted comparison ACE-R vs. MMSE for diagnosis of dementia (adapted from Larner 2013b; data from Larner 2009, 2013a)

	ACE-R	MMSE
Sensitivity (Se)	0.87	0.70
Specificity (Sp)	0.91	0.89
AUC ROC	0.94	0.91
Prevalence of dementia (π)	0.35	
Δ Sensitivity	0.17	
Δ Specificity	0.02	
Weighted comparison (WC)	0.17 = net benefit	
Equivalent increase (EI)	+61	

Table 6.2 Weighted comparison MoCA vs. MMSE for diagnosis of any cognitive impairment (adapted from Larner 2013b; data from Larner 2012b)

	MoCA	MMSE
Sensitivity (Se)	0.97	0.65
Specificity (Sp)	0.60	0.89
AUC ROC	0.91	0.83
Prevalence of cognitive impairment (π)	0.43	
Δ Sensitivity	0.32	
Δ Specificity	-0.29	
Weighted comparison (WC)	0.28 = net benefit	
Equivalent increase (EI)	+121	

Table 6.3 Weighted comparison TYM vs. MMSE for diagnosis of dementia (adapted from Lerner 2013b; data from Hancock and Lerner 2011)

	TYM	MMSE
Sensitivity (Se)	0.73	0.79
Specificity (Sp)	0.88	0.95
AUC ROC	0.89	0.94
Prevalence of dementia (π)	0.35	
Δ Sensitivity	-0.06	
Δ Specificity	-0.07	
Weighted comparison (WC)	-0.07 = net loss	
Equivalent increase (EI)	-26	

Table 6.4 Weighted comparison MMP vs. MMSE for diagnosis of dementia (adapted from Lerner 2013b; data from Lerner 2012c)

	MMP	MMSE
Sensitivity (Se)	0.51	0.45
Specificity (Sp)	0.97	0.98
AUC ROC	0.89	0.87
Prevalence of dementia (π)	0.23	
Δ Sensitivity	0.06	
Δ Specificity	-0.01	
Weighted comparison (WC)	0.06 = net benefit	
Equivalent increase (EI)	+13	

Table 6.5 Weighted comparison 6CIT vs. MMSE: (a) for diagnosis of dementia vs. no dementia (n = 150); (b) for diagnosis of dementia vs. MCI (n = 65); (c) for diagnosis of MCI vs. subjective memory complaint (n = 128) (adapted from Abdel-Aziz and Lerner 2015)

	6CIT	MMSE
(a)		
Sensitivity (Se)	0.77	0.59
Specificity (Sp)	0.80	0.85
Prevalence of dementia (π)	0.147	
Δ Sensitivity	0.18	
Δ Specificity	-0.05	
Weighted comparison (WC)	0.15 = net benefit	
Equivalent increase (EI)	+22	
(b)		
Sensitivity	0.77	0.59
Specificity	0.65	0.74
Prevalence of dementia (π)	0.338	
Δ Sensitivity	0.18	
Δ Specificity	-0.09	
WC	0.16 = net benefit	
EI	+55	
(c)		
Sensitivity	0.56	0.51
Specificity	0.71	0.75
Prevalence of MCI (π)	0.336	
Δ Sensitivity	0.05	
Δ Specificity	-0.04	
WC	0.04 = net benefit	
EI	+14	

Table 6.6 Weighted comparison MACE (cut-off $\leq 25/30$) vs. MMSE (cut-off $\leq 24/30$): (a) for diagnosis of dementia vs. no dementia ($n = 135$); (b) for diagnosis of MCI vs. subjective memory complaint ($n = 111$) (adapted from Lerner 2015a)

	MACE	MMSE
(a)		
Sensitivity (Se)	1.00	0.92
Specificity (Sp)	0.28	0.72
Prevalence of dementia (π)	0.177	
Δ Sensitivity	0.08	
Δ Specificity	-0.45	
Weighted comparison (WC)	-0.13 = net loss	
Equivalent increase (EI)	-22	
(b)		
Sensitivity	1.00	0.54
Specificity	0.43	0.86
Prevalence of MCI (π)	0.35	
Δ Sensitivity	0.46	
Δ Specificity	-0.43	
WC	0.38 = net benefit	
EI	+133	

Table 6.7 Weighted comparison TYM vs. ACE-R (adapted from Hancock and Lerner 2011, corrected from Lerner 2014a:137)

	TYM	ACE-R
Sensitivity (Se)	0.73	0.90
Specificity (Sp)	0.88	0.93
AUC ROC	0.89	0.98
Prevalence of dementia (π)	0.35	
Δ Sensitivity	-0.17	
Δ Specificity	-0.05	
Weighted comparison (WC)	-0.18 = net loss	
Equivalent increase (EI)	-63	

Table 6.8 Weighted comparison MACE (cut-off $\leq 25/30$) vs. MoCA (cut-off $\geq 26/30$): (a) for diagnosis of dementia vs. no dementia ($n = 260$); (b) for diagnosis of MCI vs. subjective memory complaint ($n = 217$) (adapted from Lerner 2017a)

	MACE	MoCA
(a)		
Sensitivity (Se)	0.98	1.00
Specificity (Sp)	0.35	0.31
Prevalence of dementia (π)	0.17	
Δ Sensitivity	-0.02	
Δ Specificity	0.04	
Weighted comparison (WC)	-0.00047 = net loss	
Equivalent increase (EI)	< -1 (= -0.08)	
(b)		
Sensitivity	0.95	0.92
Specificity	0.51	0.44
Prevalence of MCI (π)	0.29	
Δ Sensitivity	0.03	
Δ Specificity	0.07	
WC	0.047 = net benefit	
EI	+14	

Table 6.9 Weighted comparison AD8 vs. (a) MMSE (n = 125), and (b) 6CIT (n = 169) for diagnosis of cognitive impairment vs. no cognitive impairment (adapted from Lerner 2015c)

(a)		
	AD8	MMSE
Sensitivity (Se)	0.97	0.53
Specificity (Sp)	0.15	0.75
Prevalence of cognitive impairment (π)	0.576	
Δ Sensitivity	0.44	
Δ Specificity	-0.60	
Weighted comparison (WC)	0.40 = net benefit	
Equivalent increase (EI)	+230	
(b)		
	AD8	6CIT
Sensitivity	0.96	0.72
Specificity	0.17	0.55
Prevalence of cognitive impairment (π)	0.621	
Δ Sensitivity	0.24	
Δ Specificity	-0.38	
WC	0.22 = net benefit	
EI	+137	

Table 6.10 Weighted comparison ACE-R vs. MMSE for diagnosis of dementia (data adapted from Lerner and Hancock 2014)

	ACE-R	MMSE
Sensitivity (Se)	0.82	0.62
Specificity (Sp)	0.89	0.95
Prevalence of dementia (π)	0.276	
Δ Sensitivity	0.20	
Δ Specificity	-0.06	
Weighted comparison (WC)	0.18 = net benefit	
Equivalent increase (EI)	+50	

The various WC and EI findings are summarised in Table 6.11 (Lerner 2015d:105). The data suggest that for the diagnosis of dementia ACE-R is superior to MMSE and TYM; for diagnosis of any cognitive impairment MoCA and AD8 are superior to MMSE, and AD8 is superior to 6CIT; and for diagnosis of MCI MACE is superior to MMSE. All WC evaluations were in the same direction as the available values for AUC ROC, i.e. favoured ACE-R, MoCA, MMP, and 6CIT vs. MMSE, favoured MMSE vs. TYM, and favoured ACE-R vs. TYM (Lerner 2013b).

The calculation of WC and EI is largely dependent on differences in test sensitivity, which are ultimately dependent on the test cut-off used, like many other measures of discrimination derived from the 2×2 table (Sect. 2.3.2). Choice of a different method for determining test cut-off may potentially change the outcome of weighted comparisons, from net benefit to net loss (Lerner 2015e).

Table 6.11 Summary of weighted comparison and equivalent increase between CSIs for identification of (a) dementia vs. no dementia, (b) any cognitive impairment (= dementia + MCI) vs. no cognitive impairment, and (c) MCI vs. subjective memory complaint (adapted from Larner 2015d:105)

	Weighted comparison (WC)	Classification	Equivalent increase (EI)
(a)			
ACE-R vs. MMSE	0.17; 0.18	Net benefit	+61; +50
TYM vs. MMSE	-0.07	Net loss	-26
MMP vs. MMSE	0.06	Net benefit	+13
6CIT vs. MMSE	0.15	Net benefit	+22
TYM vs. ACE-R	-0.18	Net loss	-63
MACE vs. MMSE	-0.13	Net loss	-22
MACE vs. MoCA	-0.00047	Net loss	< -1
(b)			
MoCA vs. MMSE	0.28	Net benefit	+121
AD8 vs. MMSE	0.40	Net benefit	+228
AD8 vs. 6CIT	0.22	Net benefit	+137
(c)			
6CIT vs. MMSE	0.04	Net benefit	+14
MACE vs. MMSE	0.38	Net benefit	+133
MACE vs. MoCA	0.047	Net benefit	+14

6.1.2 Q* Index

Another potentially useful summary measure denoting the diagnostic value of a screening instrument is the Q* index derived from the ROC curve (Walter 2002). Q* index is defined as the “point of indifference on the ROC curve”, where the sensitivity and specificity are equal, or, in other words, where the probabilities of incorrect test results are equal for disease cases and non-cases (i.e. indifference between false positive and false negative diagnostic errors, with both assumed to be of equal value/cost). The Q* index is that point in ROC space which is closest to the ideal top left-hand (“northwest”) corner of the ROC curve, where the anti-diagonal through ROC space intersects the ROC curve (Fig. 6.1).

Q* index was derived for a number of CSIs examined in pragmatic diagnostic test accuracy studies undertaken in CFC (Larner 2015f; Table 6.12). Q* index ranged from 0.88 for ACE-R to 0.76 for MACE. The ranking of Q* index for the various CSIs examined paralleled that for AUC ROC, with ACE-R ranked highest and MACE lowest using either parameter.

Comparing the Q* index cut-off point with cut-offs defined in CSI index studies, the former was always lower (and hence less sensitive but more specific) than the latter. Comparing test sensitivity and specificity at the Q* index cut-off point showed that for all CSIs with the exception of ACE-R, Q* index-derived test cut-offs lay between those derived from maximal correct classification accuracy and maximal Youden index. Hence, if Q* index point were used as the test cut-off, it was more

Fig. 6.1 (a) Typical receiver operating characteristic (ROC) curve or plot with diagonal or chance line (data for ACE-R adapted from Larner 2009, 2013b, see Fig. 4.4) reprinted with permission; (b) typical ROC curve (same data points as a) with anti-diagonal line: where the lines cross in ROC space indicates equal test sensitivity and specificity, by definition the Q* index (the point closest to the ideal top left-hand corner of the ROC curve) (Larner 2015f) reprinted with permission

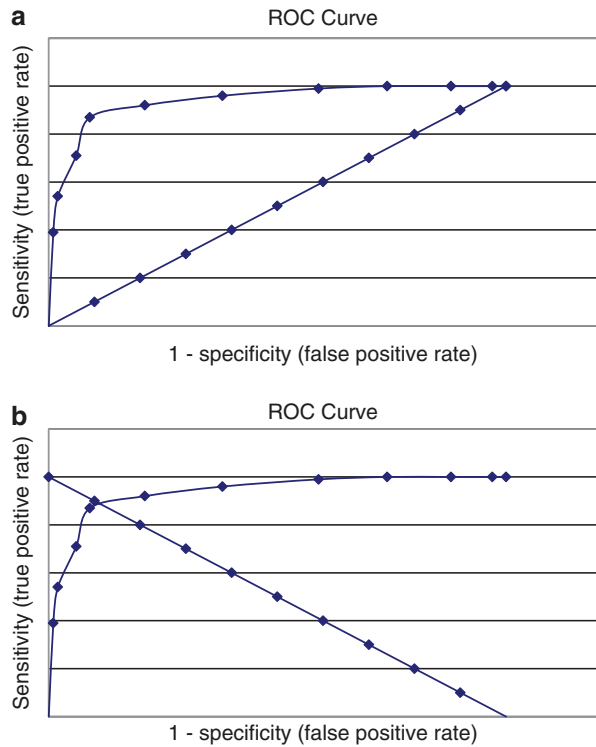


Table 6.12 Summary of Q* index for various CSIs compared with area under the receiver operating characteristic curve (AUC ROC) (adapted from Larner 2015f)

CSI	Q* index (ranking)	AUC ROC (95% CI) (ranking)	Reference
MMSE	0.82 (2)	0.91 (0.88–0.95) (2=)	Larner (2013b)
ACE-R	0.88 (1)	0.94 (0.91–0.97) (1)	Larner (2013b)
MoCA	0.79 (4)	0.91 (0.86–0.95) (2=)	Larner (2012b)
TYM	0.80 (3)	0.89 (0.84–0.93) (4)	Hancock and Larner (2011)
MACE	0.76 (5)	0.86 (0.83–0.90) (5)	Larner (2015a)

sensitive (and less specific) than if using the maximal correct classification accuracy cut-off, and less sensitive (and more specific) than if using the maximal Youden index cut-off. Q* index cut-offs reduced the sensitivity of very sensitive tests such as the ACE-R, MoCA, TYM and MACE $\leq 25/30$, but improved sensitivity for very specific tests such as MACE $\leq 21/30$ (Larner 2015f).

If a metric to compare diagnostic tests is required, Q* index has merit and, since it is based on sensitivity and specificity, may perhaps be preferred to AUC ROC results as a more intuitive measure.

6.1.3 Comparing Test Speed Versus Test Accuracy

The trade-off between speed and accuracy in the performance of voluntary movements, such that more accurate movements are performed more slowly, has long been recognised (Woodworth 1899). This speed-accuracy trade-off may perhaps apply to any task, and since speed is inversely proportional to time it may also be formulated as a time-accuracy trade-off, longer times being required for greater accuracy.

Is there is a trade-off between CSI diagnostic accuracy and administration time, or in other words are shorter CSIs less accurate than longer ones which may sample more cognitive domains and/or in greater depth? This was examined for a number of CSIs used in CFC by comparing parameters of test diagnostic accuracy against duration of test administration. The latter is not routinely measured in the clinical setting (there are exceptions when a stopwatch has been used, but this is usually for research purposes; Lees et al. 2017), although approximate timings can be given (see Sect. 2.1.3, Box 2.1). Hence, more easily accessible surrogate measures of test duration were used, namely either the overall test score or the total number of items/questions in the test (Larner 2015g, h).

Two measures of diagnostic accuracy, the correct classification accuracy or overall test accuracy (defined as the sum of true positives and true negatives divided by the total number of patients tested; Sect. 2.3.2, Box 2.3) and the area under the receiver operating characteristic curve (AUC ROC) were plotted (= output or effect, hence the dependent variable, y axis) against overall test score and against the total number of items/questions in the test (= inputs or causes, hence independent variables, x axis). Correlations between correct classification accuracy and AUC ROC and the surrogate time measures were also calculated.

Data (Table 6.13) were extracted from several pragmatic prospective diagnostic test accuracy studies examining nine performance-based CSIs: Addenbrooke's Cognitive Examination (ACE), Addenbrooke's Cognitive Examination-Revised (ACE-R), DemTect, Mini-Addenbrooke's Cognitive Examination (MACE), Mini-Mental Parkinson (MMP), Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), Six-item Cognitive Impairment Test (6CIT), and the Test Your Memory (TYM) test (see Chap. 4 for index studies).

Correct classification accuracy was positively correlated with both total test score ($r = 0.58$) and with total number of test items/questions ($r = 0.66$). Both correlations were classified as moderate, and respectively either did not reach statistical significance ($t = 1.89$, $df = 7$, $p > 0.1$) or showed a trend towards significance ($t = 2.33$, $df = 7$, $0.1 > p > 0.05$).

AUC ROC curve was positively correlated with total test score ($r = 0.83$; Fig. 6.2) and with total number of test items/questions ($r = 0.79$; Fig. 6.3). Both correlations were classified as high and both reached statistical significance ($t = 3.86$, $df = 7$, $p < 0.01$; and $t = 3.46$, $df = 7$, $p < 0.02$, respectively).

These analyses suggested that there is a trade-off for CSIs between two surrogate measures of duration of test administration and two measures of test diagnostic accuracy. Investing more time during the clinical encounter in administering longer CSIs might therefore pay dividends in terms of improved accuracy of dementia

Table 6.13 Approximate administration time for cognitive screening instruments (CSIs) and surrogate measures thereof (total test score, total number of test items/questions) with diagnostic accuracy (overall correct classification and area under ROC curve) for diagnosis of dementia (adapted from Larner 2015h)

CSI	Approximate, estimated, administration time (min)	Total test score	Number of test items or questions	Accuracy (95% CI)	AUC ROC (95% CI)	Data source
ACE	15–20	100	52	0.84 (0.80–0.88)	0.93 (0.90–0.96)	Larner (2007a)
ACE-R	15–20	100	66	0.89 (0.85–0.93)	0.94 (0.91–0.97)	Larner (2013a)
DemTect	8–10	18	13	0.78 (0.71–0.86)	0.87 (0.80–0.93)	Larner (2007b)
MACE	5–10	30	10	0.84 (0.78–0.91)	0.86 (0.83–0.90)	Larner (2015a)
MMP	5–10	32	23	0.86 (0.81–0.91)	0.89 (0.84–0.94)	Larner (2012c)
MMSE	5–10	30	21	0.86 (0.81–0.90)	0.87 (0.81–0.92)	Larner (2012c)
MoCA	10–15	30	22	0.81 (0.75–0.88)	0.91 (0.86–0.95)	Larner (2012b)
6CIT	2–3	28	7	0.80 (0.75–0.85)	0.90 (0.85–0.95)	Abdel-Aziz and Larner (2015)
TYM	5–10 (self-administered under medical supervision)	50	25	0.83 (0.78–0.88)	0.89 (0.84–0.93)	Hancock and Larner (2011)

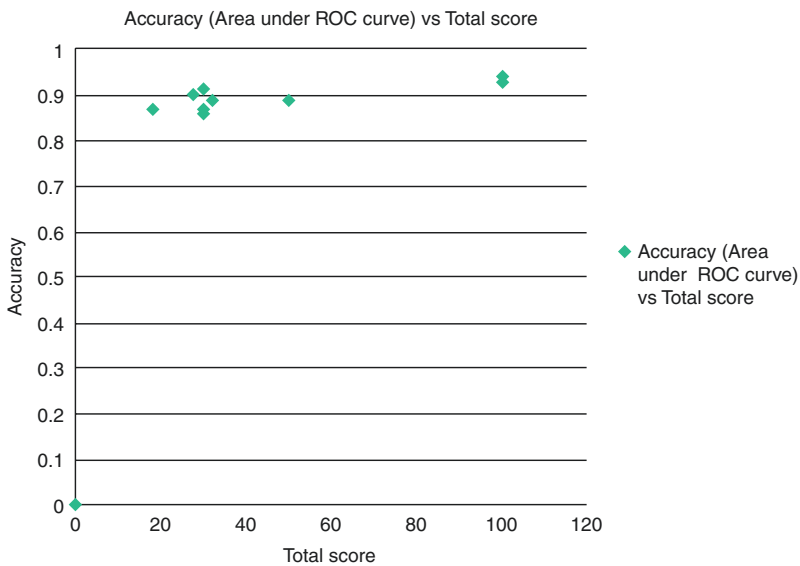


Fig. 6.2 Scatter plot of area under ROC curve (= measure of accuracy) versus total test score (= surrogate measure of test administration time) (adapted from Larner 2015h) reprinted with permission

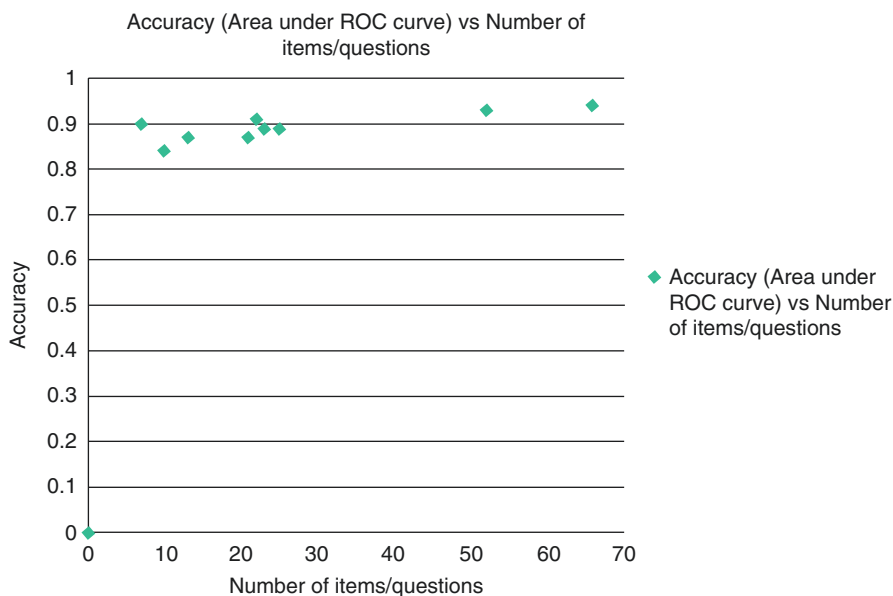


Fig. 6.3 Scatter plot of area under ROC curve (= measure of accuracy) versus total number of test items/questions (= surrogate measure of test administration time) (adapted from Larner 2015h) reprinted with permission

diagnosis. In light of these findings it might be argued on pragmatic grounds that a policy of longer outpatient clinic appointments in clinic templates (Sect. 2.1.3) for patients with cognitive complaints (e.g. 45–60 min), as compared to general neurology outpatient appointments (e.g. 15–30 min), is justified in order to permit adequate time for the administration of longer CSIs to facilitate the desired outcome of more accurate diagnosis. To borrow informally an analogy from science and engineering, there may be a lower “signal to noise ratio” when using longer CSIs (where the delivered strength of “signal” is related to statistical significance, and “noise” to standard deviation) due to their increased “bandwidth” (i.e. broader range of test scores or items) (Larner 2015h). The greater neuropsychological coverage of longer CSIs, one of the desiderata suggested by expert consensus (Malloy et al. 1997), may reduce test ceiling and floor effects.

6.1.4 Meta-Analysis: ACE and ACE-R

Meta-analysis is now a standard statistical approach to combine the results of multiple studies to improve estimates of effect size or resolve uncertainties when individual results disagree.

A meta-analysis of studies was undertaken to better understand ACE (Mathuranath et al. 2000; Sect. 4.1.5.1) and ACE-R (Mioshi et al. 2006; Sect. 4.1.5.3) utility (Larner and Mitchell 2014), using methods similar to those applied in previous meta-analyses of MMSE diagnostic accuracy (Mitchell 2009, 2013, 2017).

Literature search to end May 2013 identified 29 reports of studies of the ACE, 13 using the English version and 16 using non-English versions. All the studies identified were from high prevalence specialist secondary care settings. After application of exclusion criteria, 5 ACE studies were deemed suitable for meta-analysis (Mathuranath et al. 2000; Garcia-Caballero et al. 2006; Lerner 2007a; Stockholm et al. 2009; Yoshida et al. 2011). Across the 5 included studies there were 529 cases of dementia out of a population of 1090, a prevalence of 49%. There was no evidence of publication bias (Harbord bias = -8.23 , 95% CI = -29.1 to 12.6 , $p = 0.37$; Harbord et al. 2006).

Pooling the raw data from these studies demonstrated that 512 out of 529 cases were correctly identified using the ACE, giving a pooled sensitivity of 0.968. On meta-analytic weighting this was corrected to 0.969 (95% CI = 0.927 to 0.994). Non-cases (377) were correctly ruled-out from a sample of 561 comparison subjects to give a pooled specificity of 0.672. On meta-analysis this was corrected to 0.774 (95% CI = 0.583 to 0.918). Unadjusted the PPV was therefore 0.747 and the NPV 0.955 (Lerner and Mitchell 2014).

Literature search to end May 2013 identified 31 reports of studies of the ACE-R, 16 using the English version and 15 using non-English versions. All the studies identified were from high prevalence specialist secondary care settings. After application of exclusion criteria, 5 studies were deemed suitable for meta-analysis (Mioshi et al. 2006; Lerner 2009, 2013a; Alexopoulos et al. 2010; Yoshida et al. 2012; Dos Santos Kawata et al. 2012). Across the 5 included studies there were 560 cases of dementia out of a population of 1156, a dementia prevalence of 48%. Harbord bias was not significant (0.097, 95% CI = -18.95 to 19.14 , $p = 0.99$; Harbord et al. 2006).

Pooling the raw data from these studies demonstrated that 514 out of 560 cases were correctly identified using the ACE-R, giving a pooled sensitivity of 0.918. This was adjusted on meta-analysis to 0.957 (95% CI = 0.922 to 0.982). Non-cases (383) were correctly ruled-out from a sample of 596 comparison subjects to give a pooled specificity of 0.643. This was corrected on meta-analysis to 0.875 (95% CI = 0.638 to 0.994). Unadjusted the PPV was therefore 0.707 and the NPV 0.893 (Lerner and Mitchell 2014).

Combining the studies ($n = 9$) which used the MMSE against either the ACE ($n = 5$) or ACE-R ($n = 4$) generated a pooled MMSE sensitivity of 0.920 (95% CI = 0.849 to 0.968) and specificity of 0.869 (95% CI = 0.805 to 0.921) (Lerner and Mitchell 2014), inverting the pattern of low sensitivity and high specificity typically seen in diagnostic test accuracy studies of MMSE (see Sect. 4.1.1; Tables 4.1–4.7).

6.1.5 Effect Size (Cohen's d)

Effect size may be denoted by a variety of summary indices, of which Cohen's d is probably the most commonly used in the medical literature (Cohen 1988). This parameter is calculated as the difference of the means of two groups divided by the weighted pooled standard deviations of the groups (see Sect. 2.3.2, Fig. 2.2). Cohen

(1988, 1992) suggested that effect sizes of 0.2 to 0.3 were small, 0.5 medium, and ≥ 0.8 large.

Effect size (Cohen's *d*) for a number of the CSIs examined in CFC has been calculated (Larner 2014b, 2016b) based on data from previous pragmatic diagnostic accuracy studies which examined the MMSE, MMP, 6CIT, MoCA, TYM, ACE-R, AD8, and MACE. Mean test scores for demented and non-demented groups, and for mild cognitive impairment and subjective memory complaint groups, along with their standard deviations, were applied to the Cohen's *d* formula to calculate effect sizes.

Comparing patients with dementia and no dementia suggested large but similar effect sizes for all of the CSIs examined (Table 6.14). These values suggested a consistent difference in test scores between demented and non-demented individuals.

Comparing patients with mild cognitive impairment and no dementia (subjective memory complaint) suggested smaller effect sizes for all of the CSIs examined than in the dementia versus no dementia distinction (Table 6.15). However, effect sizes for the MoCA and MACE were larger than for other tests. These values suggested a

Table 6.14 Effect size (Cohen's *d*) for diagnosis of dementia versus no dementia (MCI + SMC) (adapted from Larner 2014b, 2015d:99)

CSI	Cohen's <i>d</i>	Study
ACE-R	1.87 (large)	Larner and Hancock (2014)
AD8	0.84 (large)	Larner (2015c)
MACE	1.52, 1.71 (large)	Larner (2015a) and Williamson and Larner (2018)
MMP	1.78 (large)	Larner (2012c)
MMSE	1.48, 1.59, 1.56 (large)	Larner (2012b, c, 2015a)
MoCA	1.80, 2.01 (large)	Larner (2012b, 2017a)
6CIT	1.89 (large)	Abdel-Aziz and Larner (2015)
TYM	1.62 (large)	Hancock and Larner (2011)

Table 6.15 Effect size (Cohen's *d*) for diagnosis of mild cognitive impairment versus subjective memory complaint (adapted from Larner 2014b, 2015d:99, 2016b)

CSI	Cohen's <i>d</i>	Study
ACE-R	0.73 (medium)	Larner and Hancock (2014)
AD8	0.31 (medium)	Larner (2015c)
MACE	1.59, 1.23 (large)	Larner (2015a) and Williamson and Larner (2018)
MMP	0.81 (large)	Larner (2012c)
MMSE	0.92 (large), 0.69 (medium), 1.26 (large)	Larner (2012b, c, 2015a)
MoCA	1.45, 1.25 (large)	Larner (2012b, 2017a)
s-MoCA	1.19, 1.37	Larner (2017b)
6CIT	0.65 (medium)	Abdel-Aziz and Larner (2015)
TYM	0.48 (medium)	Hancock and Larner (2011)

consistent difference in test scores between MCI and non-demented individuals, but with MoCA and MACE performing best. Since MoCA was designed to identify MCI cases (Nasreddine et al. 2005) this observation might be anticipated.

Looking at subgroups of older people (age ≥ 65 years) suggested larger effect sizes in this at-risk group in these cohorts (Table 6.16, Fig. 6.4; Wojtowicz and Lerner 2017).

Table 6.16 Effect size (Cohen’s d) for whole cohorts and for older (≥ 65 years) subgroups for diagnoses of dementia versus MCI and MCI versus SMC (adapted from Wojtowicz and Lerner 2017)

	Cohen’s d (Effect size)		Study
	Dementia vs. MCI	MCI vs. SMC	
MMSE-all	0.79 (medium)	1.03 (large)	Larner (2015a, b)
MMSE-old	0.88 (large)	1.41 (large)	
MACE-all	1.08 (large)	1.11 (large)	Larner (2015a, b, 2017a)
MACE-old	1.37 (large)	1.82 (large)	
MoCA-all	1.42 (large)	1.25 (large)	Larner (2017a)
MoCA-old	1.76 (large)	1.72 (large)	
6CIT-all	1.49 (large)	0.65 (medium)	Abdel-Aziz and Larner (2015)
6CIT-old	1.79 (large)	0.83 (large)	
AD8-all	0.71 (medium)	0.31 (medium)	Larner (2015c)
AD8-old	0.90 (large)	0.77 (medium)	
s-MoCA-all	1.33 (large)	1.37 (large)	Larner (2017b)
s-MoCA-old	1.46 (large)	1.70 (large)	

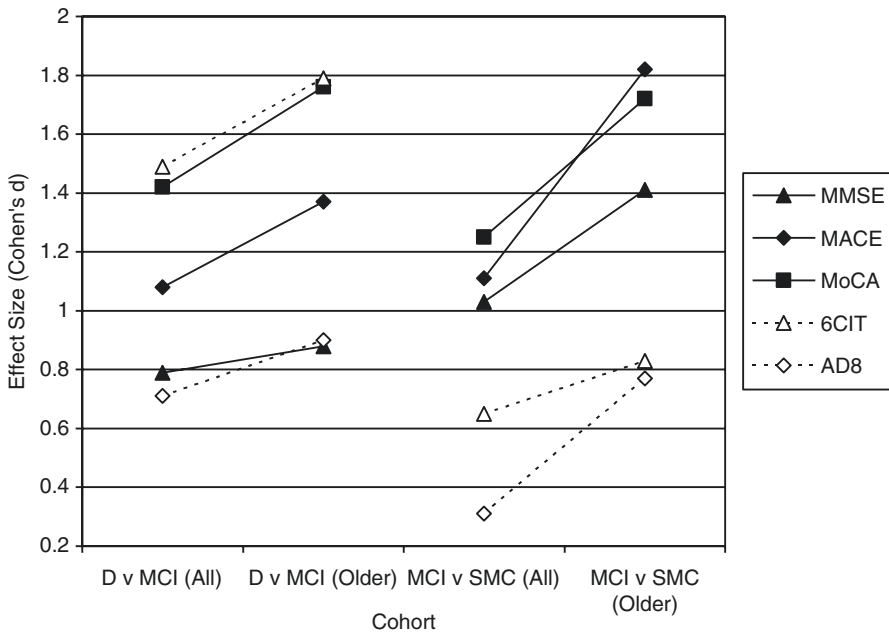


Fig. 6.4 Effect sizes (Cohen’s d) for whole cohorts and for older (≥ 65 years) subgroups for diagnoses of dementia versus MCI and MCI versus SMC (Wojtowicz and Lerner 2017) reprinted with permission

6.1.6 Correlation

Calculation of a correlation coefficient (e.g. Pearson product moment correlation coefficient) between the scores of different screening instruments applied to the same population is often performed. In diagnostic test accuracy studies correlation is not necessarily taken to imply causation (unlike research into disease aetiology, although correlation is never equivalent to causality), and hence any correlates of the target disorder may be potentially diagnostically useful, independent of any causal interpretation. Correlation is a measure of the strength of association between datasets, but it is also sometimes incorrectly assumed that high correlations give a measure of how well tests agree. Whilst the potential of a new test may be suggested if it correlates with an existing test, indicating concurrent validity, correlation is not agreement. Indeed, high correlation may in fact mask lack of agreement (Bland and Altman 1986; see Sect. 6.1.8).

Examples of correlations between different CSI scores and between CSI scores and informant and/or non-cognitive screening instruments from studies undertaken in CFC are shown in Tables 6.17 and 6.18. Unsurprisingly CSI scores are generally highly

Table 6.17 Summary of correlation coefficients for different cognitive screening instruments examined in pragmatic diagnostic test accuracy studies in CFC (adapted from Lerner 2015d:100)

	<i>r</i>	Performance	<i>t</i>	<i>p</i>
MMSE - MMP	0.93	High	35.7	<0.001
MMSE - Codex	-0.68	Moderate	6.83	<0.001
MMSE - ACE	0.92	High	28.9	<0.001
MMSE - ACE-R	0.90	High	32.8	<0.001
MMSE - MACE	0.80	High	15.5	<0.001
MMSE - 6CIT	-0.73	High	13.0	<0.001
MMSE - DemTect	0.76	High	12.0	<0.001
MMSE - MoCA	0.85	High	19.2	<0.001
MMSE - TYM	0.81	High	19.9	<0.001
MMSE - H-TYM	0.22	Low	1.37	0.1 > <i>p</i> > 0.05
MMSE - s-MoCA	0.80	High	16.2	<0.001
6CIT - MACE	-0.81	High	6.56	<0.001
6CIT - H-TYM	-0.45	Low	2.55	<0.02
DemTect - ACE	0.79	High	12.5	<0.001
ACE-R - TYM	0.86	High	20.0	<0.001
MACE - MoCA	0.83	High	24.2	<0.001
MACE - s-MoCA	0.79	High	20.4	<0.001
MACE - Free-Cog ^a	0.91	High	9.25	<0.001
MoCA - s-MoCA	0.95	High	48.0	<0.001

MMSE Mini-Mental State Examination; *MMP* Mini-Mental Parkinson; *ACE* Addenbrooke's Cognitive Examination; *ACE-R* Addenbrooke's Cognitive Examination-Revised; *MACE* Mini-Addenbrooke's Cognitive Examination; *6CIT* Six-Item Cognitive Impairment Test; *MoCA* Montreal Cognitive Assessment; *TYM* Test Your Memory (TYM) test; *H-TYM* Hard Test Your Memory (TYM) test; *s-MoCA* short Montreal Cognitive Assessment

^aPreliminary data

Table 6.18 Summary of correlation coefficients for different cognitive and informant and/or non-cognitive screening instruments examined in pragmatic diagnostic test accuracy studies in CFC (adapted from Lerner 2015d:101)

	<i>r</i>	Performance	<i>t</i>	<i>p</i>
MMSE - IQCODE	-0.37	Low	4.49	< 0.001
MMSE - PHQ-9	0.01	Very low	0.08	> 0.5
MMSE - FCS	-0.10	Very low	0.45	> 0.5
MMSE - CSDD	0.12	Very low	1.85	0.1 > <i>p</i> > 0.05
MMSE - AD8	-0.23	Very low	2.62	≈ 0.01
MMSE - ZBI (full)	0.017	Very low	0.10	> 0.5
ACE-R - IADL	0.58	Moderate	6.25	< 0.001
ACE-R - IQCODE	-0.46	Low	5.46	< 0.001
ACE-R - PHQ-9	0.12	Very low	1.19	> 0.1
ACE-R - CSDD	0.26	Very low	3.62	< 0.001
MACE - ZBI (full)	-0.008	Very low	0.047	> 0.5
6CIT - AD8	0.37	Low	5.08	< 0.001

MMSE Mini-Mental State Examination; *IQCODE* Informant Questionnaire on Cognitive Decline in the Elderly; *PHQ-9* Patient Health Questionnaire-9; *FCS* Fluctuations Composite Scale; *CSDD* Cornell Scale for Depression in Dementia; *ZBI* Zarit Burden Interview; *ACE-R* Addenbrooke's Cognitive Examination-Revised; *IADL* Instrumental Activities of Daily Living (IADL) Scale; *MACE* Mini-Addenbrooke's Cognitive Examination; *6CIT* Six-Item Cognitive Impairment Test

correlated, indicating concurrent validity, whereas CSI and non-CSI scores are generally less well correlated, indicating that these tests may examine different constructs.

6.1.7 Cohen's Kappa Statistic: Test of Agreement

The “test of agreement” or Cohen's kappa statistic (Cohen 1960) compares observed diagnostic agreement with that expected by chance alone (i.e. chance corrected agreement; see Sect. 2.3.3). This metric has sometimes been used to compare diagnostic tests (Table 6.19), although it is a measure of precision rather than of accuracy.

6.1.8 Bland-Altman Limits of Agreement

As previously mentioned (Sect. 6.1.6), correlation between test scores may indicate concurrent validity, but correlation is not agreement and indeed high correlation may actually mask lack of agreement. Bland and Altman (1986) suggested a method which provides a measure of agreement between tests by estimating how far apart the two values are on average and putting an interval around this (see Sect. 2.3.3). The limits of agreement thus defined indicate how closely two methods agree, but what is accepted as “close” remains a clinical rather than a statistical judgement. The Bland Altman methodology is a simple way to evaluate bias between mean differences which avoids the potentially erroneous conclusions based on correlation analyses.

Table 6.19 Summary of Cohen's kappa statistic (test of agreement) for different cognitive screening instruments examined in pragmatic diagnostic test accuracy studies (adapted from Lerner 2015d:102)

	κ	Agreement
MMSE - ACE-R	0.72 (0.63–0.81)	Substantial
MMSE - MACE	0.44 (0.29–0.59)	Moderate
MMSE - 6CIT	0.47 (0.29–0.63)	Moderate
MMSE - MoCA	0.39 (0.26–0.53)	Fair
MMSE - TYM	0.69 (0.58–0.80)	Substantial
MMSE - IQCODE	0.23 (0.07–0.39)	Fair
MMSE - AD8	−0.05 (−0.20–0.10)	None
ACE-R - TYM	0.69 (0.56–0.83)	Substantial
ACE-R - IQCODE	0.29 (0.11–0.46)	Fair
ACE-R - IADL	0.38 (0.18–0.58)	Fair
6CIT - AD8	0.10 (−0.08–0.28)	Slight

MMSE Mini-Mental State Examination; *ACE-R* Addenbrooke's Cognitive Examination-Revised; *M-ACE* Mini-Addenbrooke's Cognitive Examination; *6CIT* Six-Item Cognitive Impairment Test; *MoCA* Montreal Cognitive Assessment; *TYM* Test Your Memory (TYM) test; *IQCODE* Informant Questionnaire on Cognitive Decline in the Elderly; *IADL* Instrumental Activities of Daily Living (IADL) Scale

Table 6.20 Limits of agreement (with 95% confidence intervals) and Pearson's product moment correlation coefficients (r) for different cognitive screening instruments (adapted from Lerner 2016c)

	n	Mean difference (d)	Standard deviation of difference (s)	Limits of agreement (d \pm 2s)	r
MMSE-MoCA	147	3.61 (3.15–4.07)	2.83	−2.05 (−2.85 to 0.59) to 9.28 (8.48 to 10.1)	0.85
MMSE-MACE	244	4.00 (3.52–4.47)	3.77	−3.55 (−4.37 to −2.72) to 11.54 (10.7 to 12.4)	0.81
MACE-MoCA	193	0.61 (0.20–1.03)	2.92	−5.23 (−5.95 to −4.51) to 6.45 (5.73 to 7.17)	0.86

MMSE Mini-Mental State Examination; *MoCA* Montreal Cognitive Assessment; *MACE* Mini-Addenbrooke's Cognitive Examination

Bland Altman methodology was used to calculate limits of agreement for three brief CSIs (MMSE, MoCA, MACE) which were contrasted with Pearson product moment correlation coefficients between test scores (Lerner 2016c). Mean differences between test scores were small (<1 for MACE versus MoCA, up to 4 for MMSE versus MACE) but the calculated limits of agreement were broad (>10 points for MMSE versus MoCA and MACE versus MoCA; and >15 points for MMSE versus MACE). Test scores were highly correlated ($r > 0.8$) in all the studies (Table 6.20). Bland-Altman plot of difference against mean for the comparison of MMSE versus MACE is shown in Fig. 6.5.

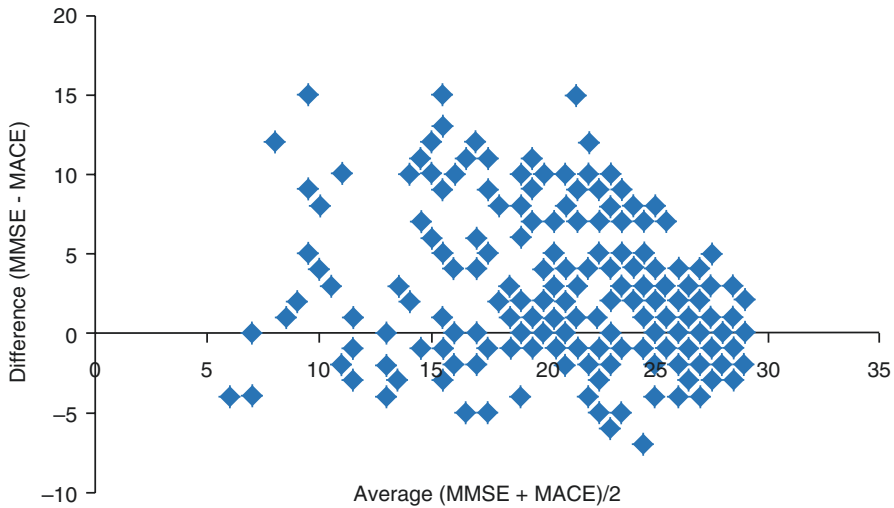


Fig. 6.5 Bland-Altman plot of difference against mean for MMSE versus MACE (Larner 2016c; data from Larner 2015a, b; $n = 244$) reprinted with permission

6.2 Combining Screening Instruments

The expectation that a single screening instrument might be entirely adequate for the diagnosis of a multidimensional construct such as the dementia syndrome, with the changes in symptomatology which occur in that syndrome over time, is likely to be wishful thinking. Different methods in staging dementia are recognised to give different results, with moderate to fair correlation of clinical scales and MMSE but a much greater dispersion of functional capacity as measured by the Instrumental Activities of Daily Living (IADL) Scale, indicating that factors other than dementia severity influence functional capacity (Juva et al. 1994). Hence combinations of tests, perhaps addressing the different domains (cognitive; functional, behavioural, global; see Chaps. 4 and 5 respectively) might be desirable, as may combinations of patient and informant information. Combinations of test results have been examined on occasion and found to give “added value” in some instances (e.g. Mackinnon et al. 2003; De Lepeleire et al. 2005).

As previously mentioned (see Sect. 2.3.2), when using screening instruments there is always a balance or trade-off to be struck between test sensitivity and specificity, with the chosen test cut-off being determined by the needs of the particular clinical situation. To optimise this trade-off, combinations of tests may be required. For example, ACE VLOM ratio showed poor sensitivity but

good specificity for the diagnosis of FTLD (see Sect. 4.1.5.2), principally because cases of bvFTD were missed (Bier et al. 2004), so combination with the Frontal Assessment Battery (FAB; see Sect. 4.2.1), which is highly sensitive for bvFTD, might be appropriate. FAB may therefore be useful as a situation-specific clinical assessment when a diagnosis of bvFTD is being considered. Use of the semantic index subscore of the ACE is appropriate if semantic dementia is being considered in the differential diagnosis (Sect. 4.1.5.2). Studies in CFC have not encouraged the view that the Ala subscore is useful prospectively for the diagnosis of DLB (Sect. 4.1.1.1), likewise the modified Ala (Sect. 4.1.5.2) and MoCA Ala (Sect. 4.1.8.1). The Mayo Fluctuations Questionnaire might be considered if DLB or PDD enters in the differential diagnosis (Ferman et al. 2004; Lerner 2012d; see Sect. 5.4.3).

Following the methodology of Flicker et al. (1997), tests may be combined either in series (both tests required to be positive before a diagnosis of dementia is made: the “And” rule) or in parallel (either test positive sufficient for a diagnosis of dementia to be made: “Or” rule); in other words, respectively, sequency and simultaneity.

6.2.1 Combining Cognitive Screening Instruments: MMSE and MoCA

The combination of the MMSE and the Clock Drawing Test (“Mini-clock”) has been reported to improve detection of mild AD and MCI (Cacho et al. 2010). Since MMSE (Folstein et al. 1975) has high specificity (see Sect. 4.1.1) and the Montreal Cognitive Assessment (MoCA; Nasreddine et al. 2005) has high sensitivity (see Sect. 4.1.8) for dementia diagnosis, the effect of combining these two cognitive screening instruments has been investigated (Lerner 2012b).

In patients administered both MoCA and MMSE ($n = 148$), combining the tests in series (“And” rule) gave results almost identical to those using the MMSE alone, whilst combining tests in parallel (“Or” rule) gave results almost identical to those using the MoCA alone (Table 6.21; compare with Tables 4.23 and 4.24). In other words, MoCA “and” MMSE was less sensitive, missing a significant proportion of the dementia and MCI cases (35% of cases) but with few false positives, whereas MoCA “or” MMSE identified almost all the cases of dementia and MCI but with a large number of false positives (greater sensitivity).

The combination of these cognitive screening instruments therefore seems to offer little over and above their individual use (Lerner 2012b). An item analysis of the MoCA and the MMSE (Damian et al. 2011) indicated that not all subtests were of equal predictive value, and that a selection of MoCA and MMSE items with high predictive value might engender a more useful hybrid test, although to the author’s knowledge this has yet to be examined.

Table 6.21 Diagnostic parameters for MMSE + MoCA in both series and parallel paradigms for diagnosis of any cognitive impairment (dementia + MCI) vs. no cognitive impairment (adapted from Lerner 2012b)

	MoCA	
<i>N</i>	150	
F:M (% female)	57:93 (38)	
Age range (years)	20–87 (median 61)	
Prevalence of cognitive impairment (dementia + MCI) (= pre-test probability)	0.43 (0.24 + 0.19)	
Pre-test odds = prevalence/(1 – prevalence)	0.75	
	Series: MoCA ≥ 26/30 + MMSE ≥ 26/30	Parallel: MoCA ≥ 26/30 + MMSE ≥ 26/30
Accuracy	0.80 (0.74–0.87)	0.75 (0.67–0.83)
Net reclassification improvement (NRI)	0.37	0.32
Sensitivity (Se)	0.65 (0.53–0.77)	0.97 (0.92–1.00)
Specificity (Sp)	0.92 (0.86–0.98)	0.59 (0.48–0.69)
<i>Y</i>	0.57	0.56
PPV (= post-test probability)	0.85 (0.75–0.95)	0.64 (0.54–0.73)
NPV	0.78 (0.70–0.86)	0.96 (0.91–1.00)
PSI	0.63	0.60
LR+	7.90 (3.79–16.5) = moderate	2.35 (1.82–3.04) = small
LR–	0.38 (0.18–0.79) = small	0.05 (0.04–0.07) = large
DOR	20.8 (9.97–43.2)	43.6 (33.7–56.4)
Post-test odds (= pre-test odds × LR+)	5.93	1.76
CUI+	0.56 (adequate)	0.62 (adequate)
CUI–	0.72 (good)	0.57 (adequate)

6.2.2 Combining Informant and Cognitive Screening Instruments

The Alzheimer Association has recommended the combined use of an informant interview with a performance measurement to detect dementia most efficiently (Cordell et al. 2013). Data from CFC which explore such combinations are presented here.

6.2.2.1 IQCODE and MMSE/ACE-R

The combination of an informant scale, the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE; see Sect. 5.4.1), and a cognitive scale, the MMSE (Sect. 4.1.1), has been previously reported in both community (Mackinnon et al. 2003) and clinical samples (Mackinnon and Mulligan 1998; Abreu et al. 2008;

Narasimhalu et al. 2008), some finding the combination helpful for detection of cases and non-cases (Mackinnon and Mulligan 1998; Mackinnon et al. 2003; Narasimhalu et al. 2008), others not (Abreu et al. 2008). This difference in findings may be related in part to the different casemix in these studies.

Many of the patients in the CFC/Brooker Centre IQCODE study (see Sect. 5.4.1) were administered the MMSE ($n = 132$) and/or the ACE-R ($n = 114$) at the same time that an informant completed the IQCODE (Hancock and Lerner 2009). The IQCODE and MMSE scores showed a low negative correlation ($r = -0.37$; $t = 4.49$, $df = 130$, $p < 0.001$). Using the test of agreement (Cohen's kappa statistic; Cohen 1960), $\kappa = 0.23$ (95% CI = 0.07–0.39), where 1 is perfect agreement between tests and 0 is agreement purely due to chance alone. For IQCODE and ACE-R, tests scores showed a low negative correlation ($r = -0.46$; $t = 5.46$, $df = 112$, $p < 0.001$) with $\kappa = 0.29$ (95% CI = 0.11–0.46).

Results of using IQCODE in combination with either MMSE ($n = 132$) or ACE-R ($n = 114$) in series or in parallel (method of Flicker et al. 1997) showed the expected improvement in specificity in the series (“And” rule) paradigm, with some reduction in sensitivity but with improved overall accuracy, PPV, diagnostic odds ratio and positive likelihood ratio (Table 6.22). There was little difference between results combining IQCODE and MMSE versus IQCODE and ACE-R, with a marginal advantage for ACE-R. In the parallel (“Or” rule) paradigm, there was the expected improvement in sensitivity, but with no change in accuracy, specificity or PPV (Hancock and Lerner 2009).

Table 6.22 Measures of discrimination for diagnosis of dementia for IQCODE, MMSE, ACE-R, IQCODE + MMSE in series or parallel, and IQCODE + ACE-R in series or parallel (adapted from Hancock and Lerner 2009)

	IQCODE $\geq 3.6/5$ + MMSE $< 24/30$ In series ($n = 132$)	IQCODE $\geq 3.6/5$ + MMSE $< 24/30$ In parallel ($n = 132$)	IQCODE $\geq 3.6/5$ + ACE-R $< 73/100$ In series ($n = 114$)	IQCODE $\geq 3.6/5$ + ACE-R $< 73/100$ In parallel ($n = 114$)
Accuracy	0.75 (0.68–0.82)	0.68 (0.60–0.76)	0.77 (0.69–0.85)	0.65 (0.56–0.74)
Sensitivity (Se)	0.64 (0.51–0.78)	0.95 (0.89–0.99)	0.67 (0.55–0.79)	0.93 (0.87–0.99)
Specificity (Sp)	0.88 (0.80–0.96)	0.36 (0.23–0.48)	0.88 (0.79–0.96)	0.36 (0.23–0.48)
Y	0.52	0.31	0.55	0.29
PPV	0.87 (0.78–0.96)	0.64 (0.55–0.74)	0.85 (0.74–0.95)	0.60 (0.50–0.70)
NPV	0.67 (0.56–0.77)	0.84 (0.70–0.98)	0.72 (0.61–0.83)	0.83 (0.68–0.98)
PSI	0.54	0.48	0.57	0.43
LR+	5.43 (2.65–11.1) = moderate	1.47 (1.20–1.79) = unimportant	5.38 (2.63–11.0) = moderate	1.45 (1.18–1.78) = unimportant
LR–	0.40 (0.20–0.83) = small	0.15 (0.13–0.19) = moderate	0.37 (0.18–0.77) = small	0.19 (0.16–0.24) = moderate
DOR	13.4 (6.56–27.5)	9.53 (7.82–11.6)	14.4 (7.02–29.4)	7.50 (6.10–9.23)
CUI+	0.56 (adequate)	0.61 (adequate)	0.57 (adequate)	0.56 (adequate)
CUI–	0.59 (adequate)	0.30 (very poor)	0.63 (adequate)	0.30 (very poor)

These results were in some ways similar to those of Narasimhalu et al. (2008) who, in an Asian population with low education, found best sensitivity for combined test use with application of the “Or” rule. Overall they found that a “weighted sum” of MMSE and IQCODE produced statistically superior area under the ROC curve and specificity results.

6.2.2.2 AD8 and MMSE/6CIT/MoCA/MACE

The combination of the AD8 informant scale (Galvin et al. 2005, 2006; see Sect. 5.4.2) and a number of cognitive screening instruments has also been examined.

In the AD8 study (Larner 2015c; Table 5.10), AD8 was combined with MMSE and with 6CIT. Combining AD8 with MMSE in series (i.e. both tests required to be positive before a diagnosis of cognitive impairment is made: the “And” rule) showed the expected improvement in specificity (0.83) but with greatly reduced sensitivity (0.50), whereas in parallel (i.e. either test positive sufficient for a diagnosis of cognitive impairment to be made: “Or” rule), sensitivity was maximised (1.0) whilst specificity was very low (0.08). Combining AD8 with 6CIT in series showed reduced sensitivity (0.70) and specificity (0.13) whilst in parallel both sensitivity (0.99) and specificity (0.59) were improved (Table 6.23).

In a subsequent study (Connon and Larner 2017; Larner 2017c), AD8 was combined with either the Montreal Cognitive Assessment (MoCA) or the Mini-Addenbrooke’s Cognitive Examination (MACE).

Over a 6-month period (May–October 2016), consecutive new outpatients attending CFC accompanied by a capable informant were administered MoCA whilst the informant completed AD8. Of 46 patient-informant dyads (F:M = 19:27, 41% female; age range 32–88 years, median 64), 13 were diagnosed with dementia (DSM-IV-TR criteria; dementia prevalence = 0.28), 22 had MCI (Petersen criteria; MCI prevalence = 0.67 of non-demented); the remainder ($n = 11$) were diagnosed with subjective memory complaints (Larner 2017c).

Using test cut-offs for cognitive impairment from index studies (AD8 $\geq 2/8$; MoCA $< 26/30$), standard measures of discrimination were calculated for individual tests and for combinations of AD8 with MoCA in series and in parallel (Table 6.24). Individually both tests were highly sensitive (> 0.95) but with low specificity (all ≤ 0.45). In series combination maintained specificity for little loss of sensitivity. Conversely in parallel combination maintained sensitivity. Predictive values were ≥ 0.8 for both combinations, with predictive summary index better for parallel combinations.

In 67 patient-informant dyads seen over an 8-month period (May–December 2016; F:M = 33:34, 49% female; age range 26–88 years, median 64), the patients were administered MACE whilst the informants completed AD8 (Connon and Larner 2017). Fourteen patients were diagnosed with dementia (DSM-IV-TR criteria), 32 with MCI (Petersen criteria). Using cut-offs defined in index studies (AD8 $\geq 2/8$; MACE $\leq 25/30$), the measures of discrimination (Table 6.25) showed both instruments were very sensitive (≥ 0.98) but not specific (≤ 0.38) for diagnosis of cognitive impairment. In series (“And” rule) combination improved diagnostic specificity for little loss of sensitivity. In parallel (“Or” rule) combination

Table 6.23 Measures of discrimination for diagnosis of cognitive impairment for AD8, MMSE, 6CIT, AD8 + MMSE in series or parallel, and AD8 + 6CIT in series or parallel (adapted and corrected from Larner 2015c)

Test	AD8 ≥2/8 212	MMSE ≤24/30 125	6CIT > 4/28 169	AD8 + MMSE In series 125	AD8 + MMSE In parallel 125	AD8 + 6CIT In series 169	AD8 + 6CIT In parallel 169
Accuracy	0.67	0.62	0.66	0.64 (0.56–0.72)	0.61 (0.52–0.69)	0.49 (0.41–0.56)	0.84 (0.78–0.90)
Sens (Se)	0.97	0.53	0.72	0.50 (0.38–0.62)	1.00	0.70 (0.62–0.79)	0.99 (0.97–1.00)
Spec (Sp)	0.17	0.75	0.55	0.83 (0.73–0.93)	0.08 (0.004–0.15)	0.13 (0.04–0.21)	0.59 (0.47–0.71)
Y	0.14	0.28	0.27	0.33	0.08	–0.17	0.58
PPV	0.65	0.75	0.72	0.80 (0.68–0.92)	0.60 (0.51–0.68)	0.57 (0.48–0.65)	0.80 (0.73–0.87)
NPV	0.78	0.54	0.55	0.55 (0.44–0.66)	1.00	0.21 (0.08–0.33)	0.97 (0.92–1.00)
PSI	0.43	0.29	0.27	0.35	0.60	–0.23	0.77
LR+	1.17	2.15	1.60	2.94 (1.55–5.58) = small	1.08 (1.00–1.17) = unimportant	0.81 (0.69–0.94) = unimportant	2.44 (1.81–3.28) = small
LR–	0.17	0.63	0.51	0.60 (0.32–1.14) = unimportant	0 = large	2.36 (2.02–2.76) = unimportant	0.016 (0.011–0.022) = large
DOR	6.63	3.44	3.16	4.89 (2.58–9.26)	∞	0.34 (0.29–0.40)	152 (113–204.5)
CUI+	0.63	0.39	0.52	0.40 (poor)	0.60 (adequate)	0.40 (poor)	0.79 (good)
CUI–	0.13	0.41	0.30	0.46 (poor)	0.08 (very poor)	0.03 (very poor)	0.58 (adequate)

Table 6.24 Measures of discrimination for diagnosis of cognitive impairment for AD8, MoCA, and AD8 + MoCA (n = 46) in series or parallel (adapted from Lerner 2017c)

Test Cut-off	AD8 ≥2/8	MoCA <26/30	AD8 + MoCA In series	AD8 + MoCA In parallel
Accuracy	0.78	0.87	0.85 (0.74–0.95)	0.80 (0.69–0.92)
Sens (Se)	0.97	1.00	0.97 (0.92–1.00)	1.00
Spec (Sp)	0.18	0.45	0.45 (0.16–0.75)	0.18 (0–0.41)
Y	0.15	0.45	0.42	0.18
PPV	0.79	0.85	0.85 (0.74–0.96)	0.80 (0.68–0.91)
NPV	0.67	1.00	0.83 (0.54–1.00)	1.00
PSI	0.46	0.85	0.68	0.80
LR+	1.19	2.20	1.78 (1.04–3.06) = unimportant	1.22 (0.93–1.61) = unimportant
LR–	0.16	0	0.06 (0.04–0.11) = large	0 = large
DOR	7.56	∞	28.3 (16.5–48.7)	∞
CUI+	0.77 (good)	0.85 (excellent)	0.83 (excellent)	0.80 (good)
CUI–	0.12 (very poor)	0.45 (poor)	0.38 (poor)	0.18 (very poor)

Table 6.25 Measures of discrimination for diagnosis of cognitive impairment for AD8, MACE, and AD8 + MACE (n = 67) in series or parallel (adapted from Connon and Lerner 2017)

Test Cut-off	AD8 ≥2/8	MACE ≤ 25/30	AD8 + MACE In series	AD8 + MACE In parallel
Accuracy	0.72	0.81	0.81 (0.71–0.90)	0.72 (0.61–0.82)
Sens (Se)	0.98	1.00	0.98 (0.94–1.00)	1.00
Spec (Sp)	0.14	0.38	0.43 (0.22–0.64)	0.10 (0–0.22)
Y	0.12	0.38	0.41	0.10
PPV	0.71	0.78	0.79 (0.68–0.90)	0.71 (0.60–0.82)
NPV	0.75	1.00	0.90 (0.71–1.00)	1.00
PSI	0.46	0.78	0.69	0.71
LR+	1.14	2.63	1.71 (1.18–2.49) = unimportant	1.11 (0.96–1.27) = unimportant
LR–	0.15	0	0.05 (0.03–0.07) = large	0 = large
DOR	7.50	∞	33.8 (23.2–49.0)	∞
CUI+	0.70 (good)	0.78 (good)	0.77 (good)	0.71 (good)
CUI–	0.11 (very poor)	0.38 (poor)	0.39 (poor)	0.10 (very poor)

maximised sensitivity but with poorer specificity. Series combination had better Youden index and correct classification accuracy than parallel combination. Predictive values were >0.7 for both combinations, with predictive summary index marginally better for parallel combination (Table 6.25).

The data from these studies suggested that series combination of AD8 and either MoCA or MACE may improve the balance of sensitivity and specificity for diagnosis of cognitive impairment, principally by improving diagnostic specificity in comparison to the use of individual tests.

6.2.3 Combining Functional and Cognitive Screening Instruments: IADL Scale and ACE-R; Free-Cog

The Instrumental Activities of Daily Living (IADL) Scale and its derivative, the 4-IADL score (see Sect. 5.1.1), are reported to correlate strongly with measures of cognitive function such as the MMSE (Lawton and Brody 1969; Barberger-Gateau et al. 1992; De Lepeleire et al. 2004). MCI patients with impaired IADL have a higher percentage of conversion to AD than MCI patients with preserved IADL (Chang et al. 2011). Hence a combination of functional and cognitive scales might possibly assist in dementia diagnosis.

The combination of a functional scale, IADL Scale, and a cognitive scale, ACE-R, has been examined in a subgroup of patients ($n = 79$; M:F = 34:45; dementia prevalence = 57%) from the IADL study (see Sect. 5.1.1; Hancock and Lerner 2007). Using the same IADL Scale cut-off ($\leq 13/14$) as used in that study, sensitivity and specificity for dementia diagnosis were comparable (Se = 0.91 vs. 0.87; Sp = 0.62 vs. 0.50). Using the same ACE-R cut-off ($\geq 73/100$) defined in the study of that instrument (see Sect. 4.1.5.3; Lerner 2009, 2013a), sensitivity and specificity for dementia diagnosis were comparable (Se = 0.76 vs. 0.87; Sp = 0.91 vs. 0.91). IADL Scale scores and ACE-R scores were moderately correlated ($r = 0.58$; $t = 6.25$, $df = 77$, $p < 0.001$) and the test of diagnostic agreement between the two tests was similarly moderate ($\kappa = 0.38$, 95% CI 0.18–0.58) (Lerner and Hancock 2012).

Results of using IADL in combination with ACE-R in series or in parallel (as per method of Flicker et al. 1997) showed the expected improvement in specificity in the series (“And” rule) paradigm, along with improved PPV, and positive likelihood ratio, but with loss of sensitivity, negative predictive value and negative likelihood ratio. In the parallel (“Or” rule) paradigm, there was the expected improvement in sensitivity, negative predictive value and negative likelihood ratio, but with loss of specificity, positive predictive value and positive likelihood ratio (Table 6.26). Parallel use might therefore be of possible advantage for increased sensitivity (case finding) (Lerner and Hancock 2012).

The Free-Cog scale (Sect. 4.1.10) attempts to incorporate assessment of cognition and function in a single instrument. Preliminary study showed that subscores for the cognitive function and executive function components had only low correlation ($r = 0.47$; $t = 2.24$, $df = 18$, $p < 0.05$), as might be anticipated when testing different constructs.

Table 6.26 Diagnostic parameters for IADL + ACE-R, in both series and parallel paradigms (Larner and Hancock 2012)

	IADL < 14/14 + ACE-R < 73/100 In series	IADL < 14/14 + ACE-R < 73/100 In parallel
Accuracy	0.80 (0.71–0.89)	0.81 (0.72–0.90)
Sensitivity (Se)	0.69 (0.55–0.82)	0.98 (0.93–1.00)
Specificity (Sp)	0.94 (0.86–1.00)	0.59 (0.42–0.75)
Y	0.63	0.57
PPV	0.94 (0.86–1.00)	0.76 (0.65–0.87)
NPV	0.70 (0.56–0.83)	0.95 (0.86–1.00)
PSI	0.64	0.71
LR+	11.7 (3.01–45.6) = large	2.37 (1.59–3.56) = small
LR–	0.33 (0.08–1.29) = small	0.04 (0.03–0.06) = large
DOR	35.4 (9.10–137.9)	62.9 (42.0–94.2)
CUI+	0.65 (good)	0.74 (good)
CUI–	0.65 (good)	0.56 (adequate)

6.3 Converting Cognitive Screening Instrument Scores

The Mini-Mental State Examination (MMSE) has been available for over 40 years and has come to be regarded as the benchmark against which other simple cognitive CSIs are compared. The development of more sensitive CSIs may have reduced the utility of MMSE, as may concerns about infringement of copyright (e.g. Newman and Feldman 2011; Mitchell 2013). However, MMSE test scores may still be used as the indicator or determinant for important clinical decisions in cognitively impaired patients, such as the initiation of prescription of cholinesterase inhibitors and/or memantine.

Different screening instruments measure slightly different things, based on their different item content (Chaps. 4 and 5), but these are all aspects of the construct of cognitive function. Simple methods to convert test scores from one of the commonly administered CSIs to another might therefore be of clinical utility.

One method to do this involves deriving a conversion table of equivalent scores from equipercenile equating with log-linear smoothing (e.g. for MMSE and MoCA: Roalf et al. 2013; van Steenoven et al. 2014). Another method is the calculation of linear regression equations of the form $y = a + bx$. For example, Kalbe et al. (2004) reported $MMSE = 19.997 + 0.567DemTect$ (other examples: for MMSE and ADAS-Cog, see Doraiswamy et al. 1997; for MMSE and one version of the clock drawing test, see Shua-Haim et al. 1997).

6.3.1 Linear Regression Equations

The datasets of several pragmatic diagnostic test accuracy studies undertaken in CFC (Abdel-Aziz and Larner 2015; Larner 2015c, 2016a, 2017a) were used to calculate regression equations of the form $y = a + bx$ (Larner 2017d), where y , the

Table 6.27 Regression equations and correlation coefficients of some commonly used cognitive screening instruments (adapted and extended from Lerner 2017d)

Compared CSIs (y vs. x)	N	Regression equation (y = a + bx)	Correlation coefficient (r)
MMSE vs. MoCA	147	y = 12.8 + 0.59x	0.85
MMSE vs. MACE	244	y = 12.8 + 0.58x	0.81
MMSE vs. 6CIT	150	y = 28.1–0.44x	–0.73
MMSE vs. AD8	125	y = 26.9–0.46x	–0.23
MACE vs. MoCA	260	y = 4.12 + 0.83x	0.83
MACE vs. s-MoCA	260	y = 10.4 + 1.18x	0.79
MoCA vs. s-MoCA	260	y = 7.53 + 1.43x	0.95

dependent or outcome variable, was approximate CSI score; x, the independent or explanatory variable, was score on a different CSI with which the first CSI was being compared; and a is the intercept and b the slope or gradient (regression coefficient) of the regression equation. Pearson product moment correlation coefficients were also calculated (Table 6.27).

As anticipated, since MoCA and MACE are scored positively and correlate positively with MMSE scores their regression coefficients with MMSE were positive, whereas for 6CIT and AD8, which are negatively scored and correlate negatively with MMSE scores, the slope of the regression line was negative, indicating lower MMSE scores for subjects with higher 6CIT and AD8 scores. Since MoCA, MACE, 6CIT and AD8 were all more sensitive than MMSE in the base studies, the intercept values of the regression equations were all high, indicating that many correct answers may be achieved on MMSE whilst the other tests remain at floor. MMSE is recognized to include relatively easy items which are of little value in patient assessment (Sect. 4.1.1). Greater coincidence of the various test scores occurred around ceiling.

Calculation and application of these regression equations is a relatively simple way to obtain approximate scores when converting between screening instruments (calculations can be easily done on a mobile phone calculator). Whether this approach might also be used outside of the secondary care clinic setting, whence the original data were generated, remains to be addressed. The regression equations derived here may be a simple way to generate approximate MMSE scores which may be used to inform clinical decision making without recourse to administering the MMSE per se and any potential copyright issues.

6.4 Summary and Recommendations

The various comparative metrics examined here suggest that a number of CSIs are suitable for the diagnosis of dementia. In the previous edition (Lerner 2014a:140), ACE-R was noted to be at or near the top in most categories, so was recommended as eminently suitable for those requiring cognitive screening in a dedicated Cognitive Function Clinic (i.e. a high prevalence setting). The withdrawal of ACE-R because

of issues around MMSE copyright was regretted; it was hoped that ACE-III (Hsieh et al. 2013) might be a suitable replacement. Other options include MoCA and MACE, both of which appear to be highly acceptable, and certainly seem to be best for diagnosis of MCI. MMSE may still retain a place, acknowledging its shortcomings, in terms of both its neuropsychological limitations and questionable ecological validity (Larner 2007c). However, MMSE is certainly not good for identification of MCI, so if this frames the clinical question then MoCA or MACE are preferable.

Combinations of CSIs with informant scales or with functional instruments may have added diagnostic value compared to CSIs in isolation, and certainly pragmatic value in planning clinical interventions (Chap. 5). Conversion between test scores may also be useful; for example, if therapeutic decision making is to be based on MMSE scores, then conversion of other CSI scores to approximate MMSE scores by using linear regression equations might be used.

References

- Abdel-Aziz K, Larner AJ. Six-item Cognitive Impairment Test (6CIT): pragmatic diagnostic accuracy study for dementia and MCI. *Int Psychogeriatr*. 2015;27:991–7.
- Abreu ID, Nunes PV, Diniz BS, Forlenza OV. Combining functional scales and cognitive tests in screening for mild cognitive impairment at a university-based memory clinic in Brazil. *Rev Bras Psiquiatr*. 2008;30:346–9.
- Alexopoulos P, Ebert A, Richter-Schmidinger T, et al. Validation of the German revised Addenbrooke's cognitive examination for detecting mild cognitive impairment, mild dementia in Alzheimer's disease and frontotemporal lobar degeneration. *Dement Geriatr Cogn Disord*. 2010;29:448–56.
- Ashford JW. Screening for memory disorders, dementia and Alzheimer's disease. *Aging Health*. 2008;4:399–432.
- Barberger-Gateau P, Commenges D, Gagnon M, Letenneur L, Sauvel C, Dartigues JF. Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. *J Am Geriatr Soc*. 1992;40:1129–34.
- Bier JC, Ventura M, Donckels V, et al. Is the Addenbrooke's Cognitive Examination effective to detect frontotemporal dementia? *J Neurol*. 2004;251:428–31.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–10.
- Cacho J, Benito-Leon J, Garcia-Garcia R, Fernandez-Calvo B, Vicente-Villardón JL, Mitchell AJ. Does the combination of the MMSE and Clock Drawing Test (Mini-clock) improve the detection of mild Alzheimer's disease and mild cognitive impairment? *J Alzheimers Dis*. 2010;22:889–96.
- Chang YL, Bondi MW, McEvoy LK, et al. Global clinical dementia rating of 0.5 in MCI masks variability related to level of function. *Neurology*. 2011;76:652–9.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
- Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
- Cohen J. A power primer. *Psychol Bull*. 1992;112:155–9.
- Connors P, Larner AJ. Combining informant (AD8) and patient (MACE) cognitive screening. Poster P0030, Association of British Neurologists Annual Meeting, Liverpool, 3–5 May, 2017.
- Cordell CB, Borson S, Boustani M, et al. Medicare Detection of Cognitive Impairment Workgroup. Alzheimer's Association recommendations for operationalizing the detection of cognitive

- impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimers Dement*. 2013;9:141–50.
- Damian AM, Jacobson SA, Hentz JG, et al. The Montreal Cognitive Assessment and the Mini-Mental State Examination as screening instruments for cognitive impairment: item analyses and threshold scores. *Dement Geriatr Cogn Disord*. 2011;31:126–31.
- De Lepeleire J, Aertgeerts B, Umbach I, et al. The diagnostic value of IADL evaluation in the detection of dementia in general practice. *Aging Ment Health*. 2004;8:52–7.
- De Lepeleire J, Heyrman J, Baro F, Buntinx F. A combination of tests for the diagnosis of dementia had a significant diagnostic value. *J Clin Epidemiol*. 2005;58:217–25.
- Doraswamy PM, Bieber F, Kaiser L, Krishnan KR, Reuning-Scherer J, Gulanski B. The Alzheimer's Disease Assessment Scale: patterns and predictors of baseline cognitive performance in multicenter Alzheimer's disease trials. *Neurology*. 1997;48:1511–7.
- Dos Santos Kawata KH, Hashimoto R, Nishio Y, et al. A validation study of the Japanese version of the Addenbrooke's Cognitive Examination-Revised. *Dement Geriatr Cogn Dis Extra*. 2012;2:29–37.
- Ferman TJ, Smith GE, Boeve BF, et al. DLB fluctuations: specific features that reliably differentiate DLB from AD and normal aging. *Neurology*. 2004;62:181–7.
- Flicker L, Logiudice D, Carlin JB, Ames D. The predictive value of dementia screening instruments in clinical populations. *Int J Geriatr Psychiatry*. 1997;12:203–9.
- 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;12:189–98.
- Galvin JE, Roe CM, Powlishta KK, et al. The AD8. A brief informant interview to detect dementia. *Neurology*. 2005;65:559–64.
- Galvin JE, Roe CM, Xiong C, Morris JE. Validity and reliability of the AD8 informant interview in dementia. *Neurology*. 2006;67:1942–8.
- Garcia-Caballero A, Garcia-Lado I, Gonzalez-Hermida J, et al. Validation of the Spanish version of the Addenbrooke's Cognitive Examination in a rural community in Spain. *Int J Geriatr Psychiatry*. 2006;21:239–45.
- Hancock P, Larner AJ. The diagnosis of dementia: diagnostic accuracy of an instrument measuring activities of daily living in a clinic-based population. *Dement Geriatr Cogn Disord*. 2007;23:133–9.
- Hancock P, Larner AJ. Diagnostic utility of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) and its combination with the Addenbrooke's Cognitive Examination-Revised (ACE-R) in a memory clinic-based population. *Int Psychogeriatr*. 2009;21:526–30.
- Hancock P, Larner AJ. Test Your Memory (TYM) test: diagnostic utility in a memory clinic population. *Int J Geriatr Psychiatry*. 2011;26:976–80.
- Hancock P, Larner AJ. Cornell Scale for Depression in Dementia: clinical utility in a memory clinic. *Int J Psychiatry Clin Pract*. 2015;19:71–4.
- Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med*. 2006;25:3443–57.
- Hsieh S, Schubert S, Hoon C, Mioshi E, Hodges JR. Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2013;36:242–50.
- Juva K, Sulkava R, Erkinjuntti T, Ylikoski R, Valvanee J, Tilvis R. Staging the severity of dementia: comparison of clinical (CDR, DSM-III-R), functional (ADL, IADL) and cognitive (MMSE) scales. *Acta Neurol Scand*. 1994;90:293–8.
- Kalbe E, Kessler J, Calabrese P, et al. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int J Geriatr Psychiatry*. 2004;19:136–43.
- Larner AJ. An audit of the Addenbrooke's Cognitive Examination (ACE) in clinical practice. *Int J Geriatr Psychiatry*. 2005;20:593–4.
- Larner AJ. Addenbrooke's Cognitive Examination (ACE) for the diagnosis and differential diagnosis of dementia. *Clin Neurol Neurosurg*. 2007a;109:491–4.

- Larner AJ. DemTect: 1-year experience of a neuropsychological screening test for dementia. *Age Ageing*. 2007b;36:326–7.
- Larner AJ. New tests are needed. *BMA News*. 2007c;19 May:10.
- Larner AJ. ACE-R: cross-sectional and longitudinal use for cognitive assessment. In: Fisher A, Hanin I, editors. *New trends in Alzheimer and Parkinson related disorders: ADPD 2009*. Collection of selected free papers from the 9th International Conference on Alzheimer's and Parkinson's disease AD/PD. Prague, Czech Republic, March 11–15, 2009. Bologna: Medimond International Proceedings; 2009. p. 103–7.
- Larner AJ. *Dementia in clinical practice: a neurological perspective*. Studies in the dementia clinic. London: Springer; 2012a.
- Larner AJ. Screening utility of the Montreal Cognitive Assessment (MoCA): in place of—or as well as—the MMSE? *Int Psychogeriatr*. 2012b;24:391–6.
- Larner AJ. Mini-Mental Parkinson (MMP) as a dementia screening test: comparison with the Mini-Mental State Examination (MMSE). *Curr Aging Sci*. 2012c;5:136–9.
- Larner AJ. Can the informant Fluctuation Composite Score help in the diagnosis of synucleinopathies? A pragmatic study. *Int J Geriatr Psychiatry*. 2012d;27:1094–5.
- 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*. 2013a;28:547–8.
- Larner AJ. Comparing diagnostic accuracy of cognitive screening instruments: a weighted comparison approach. *Dement Geriatr Cogn Disord Extra*. 2013b;3:60–5.
- Larner AJ. *Dementia in clinical practice: a neurological perspective*. Pragmatic studies in the cognitive function clinic. London: Springer; 2014a.
- Larner AJ. Effect size (Cohen's *d*) of cognitive screening instruments examined in pragmatic diagnostic accuracy studies. *Dement Geriatr Cogn Disord Extra*. 2014b;4:236–41.
- Larner AJ. Mini-Addenbrooke's Cognitive Examination: a pragmatic diagnostic accuracy study. *Int J Geriatr Psychiatry*. 2015a;30:547–8.
- Larner AJ. Mini-Addenbrooke's Cognitive Examination diagnostic accuracy for dementia: reproducibility study. *Int J Geriatr Psychiatry*. 2015b;30:1103–4.
- Larner AJ. AD8 informant questionnaire for cognitive impairment: pragmatic diagnostic test accuracy study. *J Geriatr Psychiatry Neurol*. 2015c;28:198–202.
- Larner AJ. *Diagnostic test accuracy studies in dementia*. A pragmatic approach. London: Springer; 2015d.
- Larner AJ. Optimizing the cutoffs of cognitive screening instruments in pragmatic diagnostic accuracy studies: maximising accuracy or Youden index? *Dement Geriatr Cogn Disord*. 2015e;39:167–75.
- Larner AJ. The Q* index: a useful global measure of dementia screening test accuracy? *Dement Geriatr Cogn Disord Extra*. 2015f;5:265–70.
- Larner AJ. Speed versus accuracy in cognitive assessment when using CSIs. *Prog Neurol Psychiatry*. 2015g;19(1):21–4.
- Larner AJ. Performance-based cognitive screening instruments: an extended analysis of the time versus accuracy trade-off. *Diagnostics (Basel)*. 2015h;5:504–12.
- Larner AJ. M-ACE vs. MoCA: a weighted comparison. *Int J Geriatr Psychiatry*. 2016a;31:1089–90.
- Larner AJ. Cognitive screening instruments for the diagnosis of mild cognitive impairment. *Prog Neurol Psychiatry*. 2016b;20(2):21–6.
- Larner AJ. Correlation or limits of agreement? Applying the Bland-Altman approach to the comparison of cognitive screening instruments. *Dement Geriatr Cogn Disord*. 2016c;42:247–54.
- Larner AJ. MACE versus MoCA: equivalence or superiority? Pragmatic diagnostic test accuracy study. *Int Psychogeriatr*. 2017a;29:931–7.
- Larner AJ. Short Montreal Cognitive Assessment: validation and reproducibility. *J Geriatr Psychiatry Neurol*. 2017b;30:104–8.
- Larner AJ. Does combining an informant questionnaire with patient performance scales improve diagnostic test accuracy for cognitive impairment? *Int J Geriatr Psychiatry*. 2017c;32:466–7.

- Larner AJ. Converting cognitive screening instrument test scores to MMSE scores: regression equations. *Int J Geriatr Psychiatry*. 2017d;32:351–2.
- Larner AJ, Hancock P. Does combining cognitive and functional scales facilitate the diagnosis of dementia? *Int J Geriatr Psychiatry*. 2012;27:547–8.
- Larner AJ, Hancock P. ACE-R or MMSE? A weighted comparison. *Int J Geriatr Psychiatry*. 2014;29:767–8.
- Larner AJ, Mitchell AJ. A meta-analysis of the accuracy of the Addenbrooke's Cognitive Examination (ACE) and the Addenbrooke's Cognitive Examination-Revised (ACE-R) in the detection of dementia. *Int Psychogeriatr*. 2014;26:555–63.
- Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9:179–86.
- Lees RA, Hendry BK, Broomfield 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*. 2017;32:1072–8.
- Mackinnon A, Mulligan R. Combining cognitive testing and informant report to increase accuracy in screening for dementia. *Am J Psychiatry*. 1998;155:1529–35.
- Mackinnon A, Khalilian A, Jorm AF, Korten AE, Christensen H, Mulligan R. Improving screening accuracy for dementia in a community sample by augmenting cognitive testing with informant report. *J Clin Epidemiol*. 2003;56:358–66.
- Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. *BMJ*. 2012;344:e3999.
- Malloy PF, Cummings JL, Coffey CE, et al. Cognitive screening instruments in neuropsychiatry: a report of the Committee on Research of the American Neuropsychiatric Association. *J Neuropsychiatry Clin Neurosci*. 1997;9:189–97.
- Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*. 2000;55:1613–20.
- 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;21:1078–85.
- Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *J Psychiatr Res*. 2009;43:411–31.
- Mitchell AJ. The Mini-Mental State Examination (MMSE): an update on its diagnostic validity for cognitive disorders. In: Larner AJ, editor. *Cognitive screening instruments. A practical approach*. London: Springer; 2013. p. 15–46.
- Mitchell AJ. The Mini-Mental State Examination (MMSE): update on its diagnostic accuracy and clinical utility for cognitive disorders. In: Larner AJ, editor. *Cognitive screening instruments. A practical approach*. 2nd ed. London: Springer; 2017. p. 37–48.
- Moons KGM, 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;17:447–54.
- Narasimhalu K, Lee J, Auchus AP, Chen CP. Improving detection of dementia in Asian patients with low education: combining the Mini-Mental State Examination and the Informant Questionnaire on Cognitive Decline in the Elderly. *Dement Geriatr Cogn Disord*. 2008;25:17–22.
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53:695–9.
- Newman JC, Feldman R. Copyright and open access at the bedside. *N Engl J Med*. 2011;365:2447–9.
- Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE. Comparative accuracies of two common screening instruments for the classification of Alzheimer's disease, mild cognitive impairment and healthy aging. *Alzheimers Dement*. 2013;9:529–37.
- Shua-Haim J, Koppuzha G, Shua-Haim V, Gross J. A simple score system for clock drawing in patients with Alzheimer's disease. *Am J Alzheimers Dis Other Dement*. 1997;12:212–5.

- Stokholm J, Vogel A, Johannsen P, Waldemar G. Validation of the Danish Addenbrooke's Cognitive Examination as a screening test in a memory clinic. *Dement Geriatr Cogn Disord*. 2009;27:361–5.
- van Steenoven I, Aarsland D, Hurtig H, et al. Conversion between mini-mental state examination, Montreal Cognitive Assessment, and Dementia Rating Scale-2 scores in Parkinson's disease. *Mov Disord*. 2014;29:1809–15.
- Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med*. 2002;21:1237–56.
- Williamson J, Larner AJ. MACE for diagnosis of dementia and MCI: 3-year pragmatic diagnostic test accuracy study. *Dement Geriatr Cogn Disord*. 2018;45 in press.
- Wojtowicz A, Larner AJ. Diagnostic test accuracy of cognitive screeners in older people. *Prog Neurol Psychiatry*. 2017;21(1):17–21.
- Woodworth RS. Accuracy of voluntary movements. *Psychol Rev*. 1899;3:1–101.
- Yoshida H, Terada S, Honda H, et al. Validation of Addenbrooke's cognitive examination for detecting early dementia in a Japanese population. *Psychiatry Res*. 2011;185:211–4.
- Yoshida H, Terada S, Honda H, et al. Validation of the revised Addenbrooke's Cognitive Examination (ACE-R) for detecting mild cognitive impairment and dementia in a Japanese population. *Int Psychogeriatr*. 2012;24:28–37.