



Featured Article

Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging

David R. Roalf^{a,*}, Paul J. Moberg^{a,b,d}, Sharon X. Xie^{c,d}, David A. Wolk^{b,d}, Stephen T. Moelter^e,
Steven E. Arnold^{a,b,d}

^aDepartment of Psychiatry, Neuropsychiatry, University of Pennsylvania, Philadelphia, PA, USA

^bDepartment of Neurology, University of Pennsylvania, Philadelphia, PA, USA

^cDepartment of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA

^dAlzheimer's Disease Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

^eDepartment of Behavioral and Social Sciences, University of the Sciences, Philadelphia, PA, USA

Abstract

Background: The aim of this study was to compare the utility and diagnostic accuracy of the Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE) in the diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI) in a clinical cohort.

Methods: Three hundred twenty-one AD, 126 MCI, and 140 older adults with healthy cognition (HC) were evaluated using the MMSE, the MoCA, a standardized neuropsychologic battery according to the Consortium to Establish a Registry of Alzheimer's Disease (CERAD-NB), and an informant-based measure of functional impairment, the Dementia Severity Rating Scale (DSRS). Diagnostic accuracy and optimal cut-off scores were calculated for each measure, and a method for converting MoCA to MMSE scores is presented.

Results: The MMSE and MoCA offer reasonably good diagnostic and classification accuracy as compared with the more detailed CERAD-NB; however, as a brief cognitive screening measure, the MoCA was more sensitive and had higher classification accuracy for differentiating MCI from HC. Complementing the MMSE or the MoCA with the DSRS significantly improved diagnostic accuracy.

Conclusion: The findings support recent data indicating that the MoCA is superior to the MMSE as a global assessment tool, particularly in discerning earlier stages of cognitive decline. In addition, we found that overall diagnostic accuracy improves when the MMSE or MoCA is combined with an informant-based functional measure. Finally, we provide a reliable and easy conversion of MoCA to MMSE scores. However, the need for MCI-specific measures is still needed to increase the diagnostic specificity between AD and MCI.

© 2012 The Alzheimer's Association. All rights reserved.

Keywords:

Alzheimer's disease; Mild cognitive impairment; MMSE; MoCA; Diagnostic accuracy

1. Introduction

Alzheimer's disease (AD) is a chronic, debilitating condition causing significant disease burden and mortality in adults >65 years of age [1–4]. The health and economic impact of AD has led to a pressing need to prevent or slow

disease onset and progression, and recent research efforts have focused on the transitional period from normal cognitive aging to dementia [5]. This transition period, typically associated with mild cognitive impairment (MCI), is signified by a measurable deterioration in cognitive function that is greater than expected based on an individual's age and education but which has not meaningfully affected a person's daily functioning [6]. Despite its being a major research focus in recent years [7], establishing the diagnosis of MCI

*Corresponding author. Tel.: 215-662-7559. Fax: 215-662-7903.
E-mail address: roalf@upenn.edu

with standard neuropsychologic assessment instruments has remained challenging [8].

In the absence of MCI screening measures with established cut-off scores and confidence intervals, more extensive neuropsychologic testing is often advised to reliably differentiate MCI from healthy cognitive aging (HC) or AD, although professional availability, time, and cost often do not allow this to occur. An efficient battery used at many specialized memory clinics and Alzheimer's disease research centers is the Consortium to Establish a Registry for Alzheimer's neuropsychologic battery (CERAD-NB) [9–13]. Although this battery was developed specifically for AD, it also has been used to discriminate between MCI and HC with some success [12–14]. The latter battery is briefer than standard neuropsychologic batteries, but the CERAD-NB is still lengthy and requires specialized training for proper administration, which makes it less practical for use in the typical neurology or geriatric practices. Thus, brief, validated, practical measures that can differentiate MCI and dementia would be very useful given the increase in the proportion of elderly individuals seeking medical help for memory-related changes.

Brief screening measures, such as the Mini-Mental State Examination (MMSE) [15] and Montreal Cognitive Assessment (MoCA) [16], are easily administered with little training and have demonstrated diagnostic utility [17,18], particularly in differentiating dementia from normal cognitive aging [19,20]. Recently, the diagnostic accuracy of these two screening measures has received increased attention as the importance of differentiating AD and MCI has grown [17,21]. Both the MMSE and MoCA accurately differentiate cognitive impairment (MCI or AD) from normal cognitive aging [20], but the MoCA has shown more utility in other disorders (e.g., Parkinson's disease [22]). To our knowledge, only one study has directly compared the MoCA and the MMSE in differentiating MCI from AD and healthy aging in a large clinical cohort [19] and one study examined MCI subtypes [23]. The latter study, which included AD patients, found improved accuracy for the MoCA, but the study was limited to the amnesic subtype of MCI in a Portuguese sample. There is a dearth of research comparing the diagnostic accuracy of the MoCA to other standardized measures in a well-characterized clinical dementia cohort. Furthermore, if the MoCA is a more accurate screening measure for MCI, as shown in other neurologic disorders [24,25], then a straightforward conversion of MoCA to MMSE scores would be useful given the vast amount of clinical and research that has been collected using the MMSE.

It is doubtful that brief cognitive screening instruments alone would be sufficient to accurately distinguish mild dementia from MCI. Such a distinction is made when cognitive decline is, by definition, associated with meaningful decline in daily personal, social, or occupational functioning [26–28]. Such functional decline may not be reported by the patient and requires report from a family member or other

informant. There are many observer-based functional rating scales that together with a brief cognitive measure may significantly improve diagnostic accuracy [29–35].

The aims of this study were to: (1) compare the diagnostic accuracy of the MMSE and MoCA to a more detailed assessment of neuropsychologic performance (the CERAD-NB); (2) measure the improvement in diagnostic accuracy when the MMSE or MoCA were used in combination with a measure of functional impairment (Dementia Severity Rating Scale [35]); and (3) determine a simple and reliable algorithm for conversion of MoCA to MMSE scores in diagnosis of AD and MCI in a large, well-characterized community-dwelling cohort evaluated and diagnosed by clinicians at a specialized memory center.

2. Methods

2.1. Study population

All participants were recruited from the Penn Memory Center and Clinical Core of the University of Pennsylvania's Alzheimer's Disease Center. Participants included 321 individuals with clinical consensus diagnoses of AD, 126 individuals with MCI, and 140 HC adults (Table 1). Diagnostic assessments included history and physical and neurologic examinations conducted by experienced clinicians, including the review of neuroimaging, psychometric, and laboratory data. On the basis of all these data,

Table 1
Demographic characteristics and performance for Alzheimer's disease (AD), mild cognitive impairment (MCI), and healthy comparison (HC) participants

Characteristic	AD (n = 321)	MCI (n = 126)	HC (n = 140)
Age, years	75.69 (8.21) ^{§¶}	72.29 (8.12) [¶]	71.19 (9.20)
Age range, years	53–93	52–88	50–88
Gender (men/women)	122/199	64/62 [#]	46/94
Race (AA/As/C/M/O)	45/3/232/37/4	12/5/93/15/0	22/2/107/9/0
Education, years	13.33 (4.15) ^{§¶}	14.86 (4.20) [¶]	15.91 (3.03)
GDS*	3.07 (2.92) ^{§¶}	2.69 (2.68) [¶]	0.99 (1.64)
Ethnicity (% Latino)	14%	14%	11%
MMSE	19.04 (5.88) ^{§¶}	26.00 (3.46) [¶]	29.30 (0.87)
MoCA	13.04 (6.05) ^{§¶}	20.94 (4.50) [¶]	26.83 (2.64)
Neuropsychologic score (CERAD) [†]	44.89 (13.71) ^{§¶}	64.21 (11.66) [¶]	86.87 (9.02)
Adjusted CERAD [‡]	58.29 (13.72)	76.08 (10.93)	97.28 (8.43)
DSRS	16.01 (8.50) ^{§¶}	6.20 (4.22) [¶]	0.28 (0.69)

AA, African-American; As, Asian; C, Caucasian; M, multiple races reported; O, other (e.g., Native American); GDS, Geriatric Depression Scale; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; CERAD, Consortium to Establish a Registry for Alzheimer's Disease; DSRS, Dementia Severity Rating Scale.

*AD: n = 283; MCI: n = 119; HC: n = 137.

[†]AD: n = 275; MCI: n = 122; HC: n = 138.

[‡]Adjusted for age, education, and gender based on Chandler et al [12];

[§]AD: n = 318; MCI: n = 124; HC: n = 137.

[¶]P < .01 vs MCI.

[¶]P < .01 vs HC.

[#]P < .01, larger proportion of men when compared with AD and HC.

Table 2
Diagnostic parameters for MMSE, MoCA, CERAD-NB, and DSRS across AD, MCI, and healthy older groups

(A)	HC and AD	HC and MCI	MCI and AD
(A) CERAD total score*			
AUC ($\pm 95\%$ CI)	0.99 (0.98–1.00)	0.93 ^{‡¶} (0.90–0.96)	0.84 (0.82–0.89)
Cut-off [†]	83	89	69
Sensitivity/specificity	96/97	87/90	74/73
Youden index	0.94	0.77	0.47
Classification accuracy	0.97 [#]	0.90 ^{***}	0.76
MMSE total score			
AUC ($\pm 95\%$ CI)	0.99 (0.98–0.99)	0.85 (0.80–0.89)	0.86 (0.82–0.90)
Cut-off	28	29	25
Sensitivity/specificity	96/97	82/73	77/83
Youden index	0.93	0.55	0.60
Classification accuracy	0.93	0.70	0.81 ^{††}
MoCA total score			
AUC ($\pm 95\%$ CI)	0.99 (0.97–0.99)	0.89 [§] (0.85–0.92)	0.85 (0.81–0.89)
Cut-off	23	25	19
Sensitivity/specificity	94/96	84/79	77/80
Youden index	0.91	0.63	0.57
Classification accuracy	0.95	0.80 [#]	0.79 ^{‡‡}
(B) DSRS total score			
AUC ($\pm 95\%$ CI)	0.99 ^{‡¶} (0.99–1.00)	0.96 ^{‡¶} (0.94–0.98)	0.86 (0.82–0.90)
Cut-off	3	2	10
Sensitivity/specificity	99/98	89/93	74/78
Youden index	0.97	0.82	0.52
Classification accuracy	0.98 ^{***}	0.91 ^{***}	0.75

AUC, area under the curve; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; HC, healthy older controls; MCI, mild cognitive impairment; AD, Alzheimer's disease.

*Age-, education-, gender-adjusted score (see Chandler et al [12]).

[†]Cut-off scores are expressed as greater than or equal to.

[‡]AUC was better than MMSE ($P < .001$).

[§]AUC was marginally better than MMSE ($P = .07$).

[¶]AUC was better than MoCA AUC ($P < .005$).

[#]Classification accuracy better than MMSE.

^{**}Classification accuracy better than MoCA.

^{††}Classification accuracy better than DSRS.

^{‡‡}Classification accuracy marginally better than DSRS ($P = .07$).

a consensus diagnosis was established using standardized clinical criteria for AD, MCI, or other neurologic or psychiatric conditions presenting with cognitive impairment [5,6,26,36].

Screening and cognitive assessments included the MMSE, the MoCA, and the CERAD-NB. A composite score for the CERAD battery was derived using a method described by Chandler et al [12]. We used the method from their study that excluded the MMSE from the CERAD composite total score, and also provided age-, education-, and gender-corrected scores [12]. The MoCA was administered on the same day as the clinical and neuropsychologic evaluation. In addition, the Dementia Severity Rating Scale (DSRS) was administered to assess functional capacity. The DSRS is a reliable, simple, 12-item, multiple-choice questionnaire that collects information from a knowledgeable informant on impairment severity in 12 cognitive and functional domains [37]. The MMSE, CERAD-NB, and DSRS were available during consensus diagnosis, but the MoCA was not. MCI subtypes and dementias other than AD were not analyzed in this study. HC subjects were also recruited and assessed

identically to the patients with AD and MCI. Informed consent for the use of all data was obtained from all persons, in accord with university institutional review board–approved protocols. All participants completed the MMSE and MoCA. Most, but not all, participants completed the CERAD-NB and DSRS. Most participants were also given the Geriatric Depression Scale (GDS) as a screening measure for mood disturbances [38] (see Table 1).

2.2. Statistical analysis

Demographic characteristics were compared across diagnostic groups using Pearson χ^2 or one-way analyses of variance (ANOVAs) with post hoc t -tests. The diagnosis accuracy for each measure (or combination of measures) was calculated as the area under the receiver-operating characteristic (ROC) curve (AUC). The AUC measure represents the mean sensitivity value for all possible values of specificity. A cut-off score for each measure that best differentiated diagnostic groups was determined using the Youden index [39], which maximizes the trade-offs between sensitivity

and specificity. The classification accuracy (probability of correct classification of subject with or without impairment at a given cut-off score) was calculated based on these cut-off scores (Table 2). Diagnostic accuracy of the MMSE, MoCA, and DSRS (or combination of measures) was compared with the CERAD-NB using χ^2 analysis. Classification accuracy of each measure was compared using Wilcoxon's signed rank test. These statistical analyses were performed using SAS (version 9.1) software (SAS Institute, Cary, NC).

MoCA scores were equated to MMSE scores using the equipercentile equating method [40]. Equipercentile equating has been used to equate numerous standardized tests, including equating of the Telephone Interview for Cognitive Status (TICS) scores to MMSE scores in older adults [41]. This statistical method allows for the determination of comparable test scores from two different measures on the basis of their corresponding percentile ranks. The advantage of the equipercentile equating method is that the equated scores always fall within the range of possible scores, which is not always true when using traditional mean and linear equating methods. However, this method can lead to an irregular distribution of scores, and thus a log-linear transformation [42] was used to smooth the raw scores of MoCA and MMSE into a regular distribution. Equipercentile equating and log-linear smoothing was performed using the "equate" library in the R statistical program [43–46].

3. Results

Demographic characteristics along with significance values are presented in Table 1. The diagnostic groups differed in age, education, and gender distribution, but not race or ethnicity. The groups differed in GDS score [$F(2, 536) = 30.14, P < .01$], with the AD group reporting the highest score on this measure, followed by the MCI group and the NC group. GDS mean scores were, however, <5 for all three groups and for the majority of individual participants, suggesting that no subject suffered from clinically significant depression (Table 1). The AD group performed significantly poorer than the MCI and HC groups on each task; the MCI group performed significantly poorer on each cognitive task than the HC group (Table 1). For all participants, age was weakly, but significantly, correlated with each screening and cognitive measure: age and MMSE ($r = -0.19, P < .001, n = 587$); age and MoCA ($r = -0.20, P < .001, n = 587$); age and CERAD-NB ($r = -0.25, P < .001, n = 534$); and age and DSRS ($r = 0.23, P < .001, n = 579$).

3.1. ROC analysis of standard screening measures: A comparison of CERAD-NB, MMSE, and MoCA

The ROC curve analysis was used to evaluate the diagnostic accuracy of each measure (CERAD-NB, MMSE, and MoCA) to discriminate AD and MCI from one another and from healthy cognitive aging. Graphic representations of

the ROC curves are provided in Figure 1. Table 2 (part A) shows clinically relevant cut-offs for the CERAD-NB, MMSE, and MoCA. The diagnostic accuracies of the CERAD-NB, MMSE, and MoCA were all excellent for HC vs AD, with AUCs >0.99 . Diagnostic accuracies were lower, but still very good in the CERAD-NB (0.93), MMSE (0.85), and MoCA (0.89) for HC vs MCI. The CERAD-NB resulted in a significantly higher AUC than either the MoCA [$\chi^2(1) = 6.18, P < .05$] or MMSE [$\chi^2(1) = 11.40, P < .05$], and the MoCA was nominally better at differentiating HC from MCI than the MMSE [$\chi^2(1) = 3.14, P = .07$]. Diagnostic accuracies of all three cognitive measures were the lowest when differentiating AD from MCI, yet the AUCs were still good and equally high for the three measures: CERAD-NB (0.85); MMSE (0.86); and MoCA (0.85).

3.2. Cut-off scores: CERAD, MMSE, and MoCA

3.2.1. HC vs AD

A cut-off score of 83 on the adjusted CERAD-NB, 28 on the MMSE, and 23 on the MoCA yielded the highest Youden index for discriminating between HC and AD (Table 2, part A). The classification accuracy of a score of 83 on the CERAD-NB was nominally better at classifying healthy and AD individuals than the optimal MoCA score ($Z = 1.71, P = .08$) and significantly better than the optimal MMSE score ($Z = 2.53, P = .01$). The classification accuracy of the MMSE and MoCA were comparable ($Z = 1.33, P = .18$).

3.2.2. HC vs MCI

A cut-off score of 89 on the CERAD-NB, 29 on the MMSE, and 25 on the MoCA yielded the highest Youden index for discriminating between HC and MCI (Table 2, part A). The classification accuracy of the optimal CERAD-NB score was better at classifying healthy and AD individuals than either the optimal MMSE ($Z = 5.74, P < .001$) or MoCA ($Z = 2.94, P < .01$) score. Classification accuracy of the MoCA was superior to the MMSE ($Z = 3.56, P < .001$).

3.2.3. AD vs MCI

A cut-off score of 69 on the CERAD-NB, 25 on the MMSE, and 19 on the MoCA yielded the highest Youden index for discriminating between AD and MCI (Table 2, part A). The classification accuracy of the cut-off score was equally good for the CERAD-NB, the MMSE, and the MoCA.

3.3. ROC analysis of informant-based measure: The DSRS

Area under the ROC curve, the optimal cut-off score, sensitivity, specificity, and classification accuracy for the DSRS are presented in Table 2 (part B). The DSRS provided comparably good discrimination between each group as the

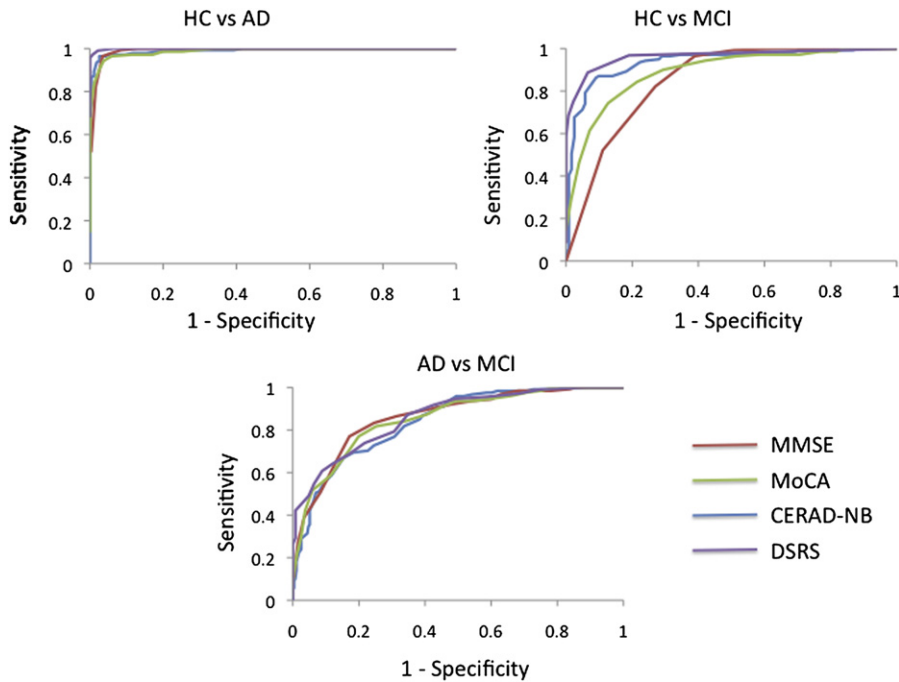


Fig. 1. ROC curve analysis of the MMSE, MoCA, CERAD-NB, and DSRS. HC, healthy controls; AD, Alzheimer's disease; MCI, mild cognitive impairment.

cognitive and screening measures. The diagnostic accuracy of the DSRS for HC and AD was equal to that of the CERAD-NB, but better than that of the MMSE [$\chi^2(1) = 6.26, P < .05$] or MoCA [$\chi^2(1) = 8.59, P < .01$], although these increases in AUC were very small ($<0.5\%$). The diagnostic accuracy of the DSRS for HC and MCI was superior to the MMSE [$\chi^2(1) = 21.67, P < .0001$] or MoCA [$\chi^2(1) = 13.52, P < .001$]. The DSRS was equally discriminant as the CERAD-NB, MMSE, and MoCA when differentiating AD and MCI.

3.4. Cut-off scores: The DSRS

A cut-off score of <3 on the DSRS yielded the highest Youden index for discriminating between HC and AD, a score of 2 was best for discriminating HC and MCI, and a score of 10 best differentiated AD from MCI. When discriminating between HC and AD or HC and MCI the DSRS yielded higher classification accuracy than the MMSE [HC vs AD: $Z = 3.89, P < .001$; HC vs MCI: $Z = 5.90, P < .001$] or MoCA [HC vs AD: $Z = 3.02, P < .01$; HC vs MCI: $Z = 3.56, P < .001$]. In contrast, the DSRS was poorer than the MMSE [$Z = 2.25, P = .03$] and nominally poorer than the MoCA [$Z = 1.84, P = .07$], but equally as accurate as the CERAD-NB when discriminating MCI and AD (Table 2, part B).

3.5. ROC analysis of combining an informant-based measure with cognitive screening tasks

Combining the DSRS with the MMSE or MoCA significantly improved diagnostic accuracy as compared

with using either screening measure alone (Table 3). Diagnostic accuracy of the MMSE improved by 13% when differentiating HC from MCI [$\chi^2(1) = 32.53, P < .0001$], and 5% when differentiating AD from MCI [$\chi^2(1) = 7.90, P < .0005$]. Using the combination of optimal cut-off scores (a score below either cut-off score) of the MMSE and/or DSRS to classify individuals resulted in an increase of 27% for HC and MCI and 12% for MCI

Table 3

Comparison of AUCs for MMSE or MoCA alone and when combined with DSRS for differentiating healthy from MCI and MCI from AD

Parameter	HC and MCI	MCI and AD
MMSE total score		
AUC	0.84 (0.80–0.89)	0.86 (0.83–0.90)
Classification accuracy	0.70	0.81
MMSE + DSRS		
AUC	0.97* (0.95–0.99)	0.91* (0.88–0.94)
Δ AUC	+0.12	+0.05
Classification accuracy	0.97 [‡]	0.93 [‡]
Δ Classification accuracy	+0.27	+0.12
MoCA		
AUC	0.88 (0.84–0.92)	0.85 (0.82–0.90)
Classification accuracy	0.80	0.79
MoCA + DSRS		
AUC	0.97 [†] (0.95–0.99)	0.90 [†] (0.87–0.93)
Δ AUC	+0.09	+0.05
Classification accuracy	0.97 [‡]	0.92 [‡]
Δ Classification accuracy	+0.17	+0.13

AUC, area under the curve; Δ AUC, change in AUC; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; HC, healthy older controls; MCI, mild cognitive impairment; AD, Alzheimer's disease.

*Score was better than MMSE AUC ($P < .0001$).

[†]Score was better than MoCA AUC ($P < .001$).

[‡]Classification accuracy higher than single measure.

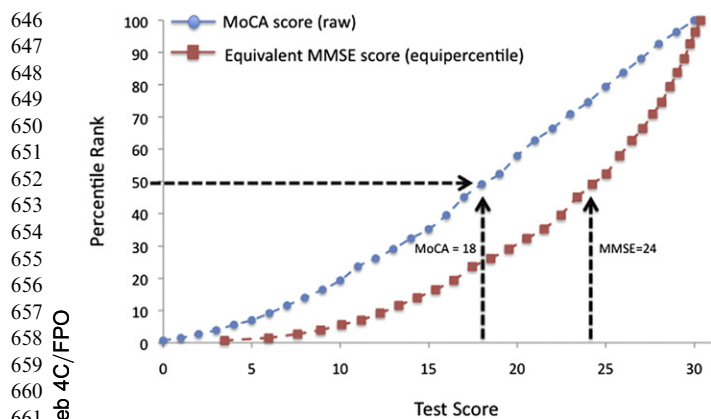


Fig. 2. Corresponding raw scores and raw percentile ranks for the MMSE and MoCA. As an example, a score of 18 on the MoCA is equivalent to a score of 24 on the MMSE.

and AD. Diagnostic accuracy (AUC) of the MoCA was improved by 9% when differentiating HC from MCI [$\chi^2(1) = 25.12, P < .001$], and 5% when differentiating AD from MCI [$\chi^2(1) = 9.84, P < .001$]. Using the combination of optimal cut-off scores of the MoCA and/or DSRS to classify individuals resulted in an increase of 17% for HC and MCI and 13% for MCI and AD. Changes in diagnostic accuracy were not assessed for HC vs AD, as diagnostic accuracy of each test alone was already excellent (>0.99).

3.6. Equating MMSE and MoCA scores

The plot of equipercentile equivalent scores on the MMSE and MoCA is presented in Figure 2. As an example, Figure 2 indicates that a score of 18 on the MoCA is equivalent to a score of 24 on the MMSE, with both of these scores falling at approximately the 50th percentile within a clinical sample with a wide range of cognitive impairment. Table 4 shows scores on the MoCA and their respective equivalents on the MMSE. In general, lower MoCA scores are equal to higher MMSE scores. For example, scores of 28–30 on the MoCA are equivalent to the highest score of 30 on the MMSE.

4. Discussion

Overall, our findings in a relatively large, community-dwelling cohort, evaluated at a specialty memory clinic, suggest that the individual screening measures (MMSE and MoCA) typically used in clinical practice to aid in the diagnosis of AD or MCI offer reasonably high classification and diagnostic accuracy when compared with a more detailed neuropsychologic evaluation (CERAD-NB). However, as a brief, stand-alone cognitive screening measure, the MoCA appears to be more sensitive than the older, but more widely used MMSE. Furthermore, using the optimal cut-off score, the classification accuracy of the MoCA exceeded that of the MMSE when differentiating MCI from HC. This finding is consistent with a recent study showing

Table 4

Conversion table for MMSE and MoCA screening measures based on equipercentile equating in 321 AD, 126 MCI, and 140 HC*

Raw MoCA score	Equivalent MMSE score
0	3
1	6
2	8
3	9
4	10
5	11
6	12
7	13
8	14
9	15
10	16
11	17
12	19
13	20
14	21
15	22
16	22
17	23
18	24
19	25
20	26
21	26
22	27
23	28
24	28
25	29
26	29
27	29
28	30
29	30
30	30

*Equivalent scores were derived from equipercentile equating with log-linear smoothing.

that the MoCA is more sensitive than the MMSE at detecting patients with mild cognitive impairment in Parkinson's disease (PD) from those with a diagnosis of PD alone. In addition, using an informant-based questionnaire of functional impairment (the DSRS) resulted in comparable, or higher, overall diagnostic accuracy for most comparisons. However, accurately differentiating MCI from AD proved the most difficult, regardless of the measure used. Finally, we have provided a simple, yet reliable, method for equating MoCA scores to traditional MMSE scores.

Our results confirm and extend prior findings on the diagnostic utility of cognitive tests in AD and MCI. First, we found that an efficient but multidimensional neuropsychologic inventory, the CERAD-NB, is more accurate at distinguishing MCI or AD from healthy individuals than brief screening measures, but not more accurate than the MMSE or MoCA at differentiating AD from MCI. The diagnostic accuracy of the CERAD-NB (93%) when differentiating MCI from the HC group in our study was very similar to that in a recent, multinational study of the diagnostic accuracy of the total CERAD score [13]. Better diagnostic accuracy using the CERAD-NB is expected as it is more

comprehensive, encompasses multiple cognitive domains, and is highly correlated with the MMSE and MoCA [12]. Thus, it can serve as a reliable standard to which the MMSE and MoCA can be compared.

We have extended upon previous findings [19,20,47,48] that the MoCA is a superior screen than the MMSE for detecting AD and MCI. Both measures easily identified AD compared with HC and were comparable with the CERAD-NB. More striking was the accuracy of the MoCA at differentiating MCI from HC as compared with the MMSE. These data suggest that the MoCA may be more sensitive to early changes in cognitive ability as it includes more robust measures of visuospatial and executive function [16]. Diagnostic accuracy declined when using these instruments to distinguish between AD and MCI. Although we found that the MMSE and MoCA were comparably accurate in differentiating AD from MCI, and the accuracy of both screens was similar to the CERAD-NB, this transition in clinical status remains a challenge for psychometric instruments. In particular, difficulty in differentiating AD from MCI is likely due to several factors including, but not limited to: (1) the heterogeneity of the MCI diagnosis; (2) the progression rates of MCI to AD of 10%–15%/year [5]; (3) limited research that has focused on cross-sectional differentiation of MCI from AD and healthy older adults (see Lonie et al [8]); and (4) the relative dearth of MCI-specific screening measures. The importance of this last point cannot be understated as the real clinical value of any test will be in its prediction of further decline, and its relationship with disease-specific biomarkers that lead to significantly improved diagnostic accuracy.

We extend upon previous findings by assessing the diagnostic accuracy of the DSRS, an informant-based measure of functional impairment. Notably, the DSRS provided better classification accuracy than the MMSE or MOCA, and comparable accuracy to the CERAD-NB, for AD and MCI when compared with HC. Yet, the diagnostic and classification accuracy of the DSRS fell, like the other measures, when differentiating AD from MCI. The DSRS may provide additional insight into the functional levels of disease severity or subtypes of MCI, especially in the absence of overt dementia. Given the capability of the DSRS to correctly discriminate dementia from healthy cognitive aging, we found that combining this measure with either the MMSE or MoCA increased diagnostic accuracy. This increase is of note as the DSRS is a reliable tool for indicating the severity of AD throughout the course of the disease [37]. Furthermore, the DSRS, unlike many cognitive measures, changes at a steady and predictable rate [35], making it an ideal measure to monitor changes in the disease course, and may provide a useful metric to measure response to treatment. Previous studies showed that combining cognitive and informant-based screens (e.g., Informant Questionnaire on Cognitive Decline in the Elderly) improved diagnostic accuracy [30,31], suggesting that measures of instrumental activities of daily living be incorporated into the diagnostic criteria for MCI [49]. To our knowledge, the pres-

ent study is the first to combine the DSRS with other screening measures to differentiate AD from MCI from HC. Given these results, further studies should consider including the DSRS during diagnostic assessment.

Finally, we used equipercentile equating to develop a conversion score between the MoCA and MMSE. Herein we have provided a table of conversion scores enables the widely recognized cut-off scores on the MMSE to be reliably compared with scores on the MoCA. Although the MMSE is the most commonly used clinical screen for monitoring acute or declining cognitive impairment, it has several limitations, such as marked ceiling effects in younger, well-educated individuals, and inconsistent performance in differentiating MCI from healthy older adults, as many individuals with MCI score above the recommended MMSE cut-off (score of 26) for impairment (for review, see Lonie et al [8]). Thus, the MMSE has limited utility in detecting subtle changes in cognition that may signal pending impairment [50,51]. The MoCA does not overcome all of the shortcomings of the MMSE, but recent data, including ours, indicate that the MoCA is a superior screening measure for classifying MCI. Equating scores on these two measures provides a straightforward way of comparing the MoCA to the MMSE, thus allowing for continuity of cognitive tracking in the clinic and comparability of data in longitudinal studies of MCI or dementing illness.

Our use of a large, well-characterized, community-dwelling cohort provides a useful perspective for determining the diagnostic accuracy of standardized global screening measures in AD and MCI. However, our study is not without limitations. The differences in gender distribution across the groups present challenges in interpreting the results. As expected, there were more women diagnosed with AD than men [52]; however, the gender distribution within the MCI group was essentially a 50/50 split. It is possible the men are overrepresented in the group due to other factors. A comparison of gender-specific diagnostic accuracy was beyond the scope of this study, but should be examined in future investigations. The impact of comorbid clinical factors, such as depression, were not considered in this study but could impact the utility of some screening measures; however, the MoCA appears robust to depression symptoms within healthy cohorts [53], and in the current sample self-reported rates of depression were low. The sample was well-educated, particularly the healthy cognition group, which makes generalization to less well-educated cohorts more challenging as education does impact MMSE [8] and MoCA [53] performance. Overall, the sample size of this study was adequate, but more precise cut-off values for each diagnostic test could be achieved with an increase in the number of MCI patients. An increase in number and subtypes of MCI, which was not considered in this study, would provide more specificity and may lead to improved diagnostic accuracy and utility. Because the MoCA shows a wider range of performances in AD and MCI than the MMSE, it seems likely that the MoCA can be used to determine differing levels of severity

or subtypes of MCI. This is particularly important as the diagnosis of MCI or AD increases with age, and thus performance on the MMSE or MoCA may systematically differ in an 80-year-old MCI patient when compared with a 60-year-old MCI patient. Thus, future studies should consider age- and domain-specific performance cut-offs.

The aim of this investigation was to compare individuals with a known diagnosis of AD or MCI to a healthy comparison group in a specialty memory and aging clinic. As a result, generalizing to other clinical settings should be done with caution. Given that the clinicians were not blinded to the MMSE, CERAD-NB, and DSRS during consensus diagnosis it is possible that diagnostic decisions were influenced by the availability of these measures. However, there was no systematic rubric employed at consensus and the data do not suggest that one or more of these measures were used preferentially to achieve diagnosis. Given the paucity of screening tools that are specific to differentiating MCI from AD and normal cognitive aging, we believe that the findings presented herein offer useful information about well-utilized standardized metrics and their diagnostic effectiveness.

In conclusion, our results support recent data that the MoCA is superior to the MMSE as a global assessment tool and we have described a reliable method for comparing the two screening measures. Furthermore, we found that when these traditional screening measures are used in the diagnosis of AD and MCI with complementary informant-based ratings of functional impairment, overall diagnostic accuracy improves.

Acknowledgments

The authors express appreciation to the research participants and staff of the Penn Memory Center/Clinical Core of the University of Pennsylvania Alzheimer's Disease Center. This work was supported by a grant from the National Institute on Aging (AG10124) the Marian S. Ware Alzheimer's Program/National Philanthropic Trust, and the National Institutes of Health (T32 MH019112).

References

- [1] Lobo A, Launer LJ, Fratiglioni L, Andersen K, Di Carlo A, Breteler MM, et al. Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. *Neurology* 2000;54(suppl 5):S4–9.
- [2] Comas-Herrera A, Northey S, Wittenberg R, Knapp M, Bhattacharyya S, Burns A. Future costs of dementia-related long-term care: Exploring future scenarios. *Int Psychogeriatr* 2011; 23:20–30.
- [3] Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, et al. Global prevalence of dementia: A Delphi consensus study. *Lancet* 2005;366:2112–7.
- [4] Wimo A, Winblad B, Aguero-Torres H, von Strauss E. The magnitude of dementia occurrence in the world. *Alzheimer Dis Assoc Disord* 2003;17:63–7.
- [5] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol* 1999;56:303–8.
- [6] Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004;256:183–94.
- [7] Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, et al. Mild cognitive impairment—beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* 2004;256:240–6.
- [8] Lonie JA, Tierney KM, Ebmeier KP. Screening for mild cognitive impairment: A systematic review. *Int J Geriatr Psychiatry* 2009; 24:902–15.
- [9] Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989;39:1159–65.
- [10] Welsh-Bohmer KA, Mohs RC. Neuropsychological assessment of Alzheimer's disease. *Neurology* 1997;49(suppl 3):S11–3.
- [11] Collie A, Shafiq-Antonacci R, Maruff P, Tyler P, Currie J. Norms and the effects of demographic variables on a neuropsychological battery for use in healthy ageing Australian populations. *Aust N Z J Psychiatry* 1999;33:568–75.
- [12] Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, et al. A total score for the CERAD neuropsychological battery. *Neurology* 2005;65:102–6.
- [13] Paajanen T, Hanninen T, Tunnard C, Mecocci P, Sobow T, Tsolaki M, et al. CERAD neuropsychological battery total score in multinational mild cognitive impairment and control populations: The AddNeuroMed study. *J Alzheimers Dis* 2010;22:1089–97.
- [14] Seo EH, Lee DY, Lee JH, Choo IH, Kim JW, Kim SG, et al. Total scores of the CERAD neuropsychological assessment battery: Validation for mild cognitive impairment and dementia patients with diverse etiologies. *Am J Geriatr Psychiatry* 2010;18:801–9.
- [15] 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.
- [16] Nasreddine ZS, Phillips NA, Bedirian 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.
- [17] Smith T, Gildeh N, Holmes C. The Montreal Cognitive Assessment: Validity and utility in a memory clinic setting. *Can J Psychiatry* 2007;52:329–32.
- [18] Damian AM, Jacobson SA, Hentz JG, Belden CM, Shill HA, Sabbagh MN, 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.
- [19] Freitas S, Simoes MR, Alves L, Santana I. Montreal Cognitive Assessment: Validation study for mild cognitive impairment and Alzheimer disease. *Alzheimer Dis Assoc Disord* 2011. Q2
- [20] Larner AJ. Screening utility of the Montreal Cognitive Assessment (MoCA): In place of—or as well as—the MMSE? *Int Psychogeriatr* 2012;24:391–6.
- [21] Lee JY, Dong Woo L, Cho SJ, Na DL, Hong Jin J, Kim SK, et al. Brief screening for mild cognitive impairment in elderly outpatient clinic: Validation of the Korean version of the Montreal Cognitive Assessment. *J Geriatr Psychiatry Neurol* 2008;21:104–10.
- [22] Dalrymple-Alford JC, McCaskill MR, Nakas CT, Livingston L, Graham C, Crucian GP, et al. The MoCA: Well-suited screen for cognitive impairment in Parkinson disease. *Neurology* 2010;75:1717–25.
- [23] Dong Y, Lee WY, Basri NA, Collinson SL, Merchant RA, Venketasubramanian N, et al. The Montreal Cognitive Assessment is superior to the Mini-Mental State Examination in detecting patients at higher risk of dementia. *Int Psychogeriatr* 2012;:1–7.
- [24] Ihara M, Okamoto Y, Takahashi R. Suitability of the Montreal Cognitive Assessment versus the Mini-Mental State Examination in detecting vascular cognitive impairment. *J Stroke Cerebrovasc Dis* 2012. Q3
- [25] MacKenzie G, Gould L, Ireland S, LeBlanc K, Sahlas D. Detecting cognitive impairment in clients with mild stroke or transient ischemic

- 1048 attack attending a stroke prevention clinic. *Can J Neurosci Nurs* 2011; 1049 33:47–50.
- 1050 [26] American Psychiatric Association. Diagnostic and statistical manual 1051 of mental disorders. 4th ed. Washington, DC: American Psychiatric 1052 Association; 2000. Text revision (DSM-IV-TR) edition.
- 1053 [27] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, 1054 Kawas CH, et al. The diagnosis of dementia due to Alzheimer's dis- 1055 ease: Recommendations from the National Institute on Aging- 1056 Alzheimer's Association workgroups on diagnostic guidelines for 1057 Alzheimer's disease. *Alzheimers Dement* 2011;7:263–9.
- 1058 [28] Jack CR Jr, Albert MS, Knopman DS, McKhann GM, Sperling RA, 1059 Carrillo MC, et al. Introduction to the recommendations from the 1060 National Institute on Aging-Alzheimer's Association workgroups on di- 1061 agnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 1062 2011;7:257–62.
- 1063 [29] Jorm AF. The Informant Questionnaire on cognitive decline in the el- 1064 derly (IQCODE): A review. *Int Psychogeriatr* 2004;16:275–93.
- 1065 [30] Flicker L, Logiudice D, Carlin JB, Ames D. The predictive value of 1066 dementia screening instruments in clinical populations. *Int J Geriatr 1067 Psychiatry* 1997;12:203–9.
- 1068 [31] Hancock P, Larner AJ. Diagnostic utility of the Informant Question- 1069 naire on Cognitive Decline in the Elderly (IQCODE) and its combina- 1070 tion with the Addenbrooke's Cognitive Examination—Revised 1071 (ACE-R) in a memory clinic-based population. *Int Psychogeriatr* 1072 2009;21:526–30.
- 1073 [32] Carr DB, Gray S, Baty J, Morris JC. The value of informant versus in- 1074 dividual's complaints of memory impairment in early dementia. *Neu- 1075 rology* 2000;55:1724–6.
- 1076 [33] Berg L. Clinical Dementia Rating (CDR). *Psychopharmacol Bull* 1077 1988;24:637–9.
- 1078 [34] Galasko D, Bennett D, Sano M, Ernesto C, Thomas R, Grundman M, 1079 et al. An inventory to assess activities of daily living for clinical trials 1080 in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. 1081 *Alzheimer Dis Assoc Disord* 1997;11(suppl 2):S33–9.
- 1082 [35] Xie SX, Ewbank DC, Chittams J, Karlawish JH, Arnold SE, Clark CM. 1083 Rate of decline in Alzheimer disease measured by a Dementia Severity 1084 Rating Scale. *Alzheimer Dis Assoc Disord* 2009;23:268–74.
- 1085 [36] Morris JC, Mohs RC, Rogers H, Fillenbaum G, Heyman A. Consor- 1086 tium to establish a registry for Alzheimer's disease (CERAD) clinical 1087 and neuropsychological assessment of Alzheimer's disease. *Psycho- 1088 pharmacol Bull* 1988;24:641–52.
- 1089 [37] Clark CM, Ewbank DC. Performance of the dementia severity rating 1090 scale: A caregiver questionnaire for rating severity in Alzheimer dis- 1091 ease. *Alzheimer Dis Assoc Disord* 1996;10:31–9.
- 1092
- 1093
- 1094
- 1095
- 1096
- 1097
- 1098
- 1099
- 1100
- 1101
- 1102
- 1103
- 1104
- 1105
- 1106
- 1107
- 1108
- 1109
- 1110
- 1111
- 1112
- 1113
- 1114
- [38] Sheikh J, Yesavage J. Geriatric Depression Scale (GDS): Recent evi- 1115 dence and development of a shorter version. *Clinical gerontology: A 1116 guide to assessment and intervention*. New York: Haworth Press; 1117 1986. p 165–73.
- [39] Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5. 1118 1119
- [40] Kolen MJ, Brennan RL. Test equating methods and practice. New 1120 York: Springer; 2005. 1121
- [41] Fong TG, Fearing MA, Jones RN, Shi P, Marcantonio ER, Rudolph JL, 1122 et al. Telephone interview for cognitive status: Creating a crosswalk 1123 with the Mini-Mental State Examination. *Alzheimers Dement* 2009; 1124 5:492–7. 1125
- [42] Moses T, vonDavier AA. An SAS macro for loglinear smoothing: Ap- 1126 plications and implications. Montreal: American Educational Re- 1127 search Association; 2005. 1128
- [43] Braun HI, Holland PW. Observed-score test equating: A mathematical 1129 analysis of some ETS equating procedures. In: Holland PW, Rubin DB, 1130 eds. Test equating. New York: Academic Press; 1982. p. 9–49. 1131
- [44] Livingston SA, Kim S. The circle-arc method for equating in small 1132 samples. *J Educ Measurement* 2009;46:330–43. 1133
- [45] Kolen MJ, Brennan RL. Test equating, scaling, and linking. 2nd ed. 1134 New York: Springer; 2004. 1135
- [46] Lord F. The standard error of equipercentile equating. *J Educ Stat* 1136 1982;7:165–74. 1137
- [47] Hancock P, Larner AJ. Test Your Memory Test: Diagnostic utility in 1138 a memory clinic population. *Int J Geriatr Psychiatry* 2011;26:976–80. 1139
- [48] Luis CA, Keegan AP, Mullan M. Cross validation of the Montreal Cog- 1140 nitive Assessment in community dwelling older adults residing in the 1141 Southeastern US. *Int J Geriatr Psychiatry* 2009;24:197–201. 1142
- [49] Brown PJ, Devanand DP, Liu X, Caccappolo E. Functional impairment 1143 in elderly patients with mild cognitive impairment and mild Alzheimer 1144 disease. *Arch Gen Psychiatry* 2011;68:617–26. 1145
- [50] Ravaglia G, Forti P, Maioli F, Servadei L, Martelli M, Brunetti N, et al. 1146 Screening for mild cognitive impairment in elderly ambulatory pa- 1147 tients with cognitive complaints. *Aging Clin Exp Res* 2005;17:374–9. 1148
- [51] Slavin MJ, Sandstrom CK, Tran TT, Doraiswamy PM, Petrella JR. 1149 Hippocampal volume and the Mini-Mental State Examination in the 1150 diagnosis of amnesic mild cognitive impairment. *AJR Am J Roent- 1151 genol* 2007;188:1404–10. 1152
- [52] Vina J, Lloret A. Why women have more Alzheimer's disease than 1153 men: gender and mitochondrial toxicity of amyloid-beta peptide. *J 1154 Alzheimers Dis* 2010;20(suppl 2):S527–33. 1155
- [53] Freitas S, Simoes MR, Alves L, Santana I. Montreal Cognitive Assess- 1156 ment: Influence of sociodemographic and health variables. *Arch Clin 1157 Neuropsychol* 2012;27:165–75. 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181