Movement Disorders
CLINICAL PRACTICE

RESEARCH ARTICLE

The Diagnostic Accuracy of Parkinson's Disease Mild Cognitive Impairment Battery Using the Movement Disorder Society Task Force Criteria

Ondrej Bezdicek, PhD,[1,] Tomas Nikolai, PhD,[1] Jiri Michalec, MA,[2] Filip Růžička, MD, PhD,[1] Petra Havránková, MD, PhD,[1] Jan Roth, MD, PhD,[1] Robert Jech, MD, PhD,[1] Evžen Růžička, MD, DSc[1]*

Abstract: **Background:** The aim of the present study was to provide empirical evidence regarding the classification accuracy of the International Parkinson and Movement Disorder Society (MDS) neuropsychological battery (NB) in the determination of Parkinson's disease mild cognitive impairment (PD-MCI).

Methods: The present cross-sectional study included 106 PD patients subjected to PD-MCI classification at Level I and 120 healthy controls (HCs). All HC and PD subjects were then assessed with MDS-NB at Level II and matched according to age and education using different thresholds (1.5 and 2.0 standard deviations [SDs] below average).

Results: We found that Level I and II resulted in different classifications of PD-MCI status. Detection thresholds of –1.5 SD and –2.0 SDs at Level II had also a significant impact on the discriminative validity of all measures in the MDS neuropsychological battery, based on area under the curve analyses. Overall, semantic fluency showed the highest potential in all comparisons not only between PD-MCI and HC, but also between PD-MCI and PD with no deficit (PD-ND).

Conclusions: Our results show that the battery at Level II is applicable and that some measures, such as semantic fluency, have high discriminative validity in the detection of PD-MCI versus PD-ND and HCs.

Parkinson's disease (PD) is a progressive neurological disorder characterized by a large number of motor and nonmotor features that can impact function to a variable degree.[1] The characteristic motor manifestations of bradykinesia, rigidity, and resting tremor are often accompanied by impairments in cognitive function.[2] Cognitive impairment in PD is heterogeneous, affecting a wide range of cognitive domains, and is progressive, with most patients developing dementia over time.[3–7]

The diagnosis of PD dementia (PDD) and predementia states, such as Parkinson's disease mild cognitive impairment (PD-MCI), has recently been established by the International Parkinson and Movement Disorder Society (MDS) Task Force.[8–10] PD-MCI was originally conceptualized as the stage between normal aging and dementia.[11–13] Patients with PD-MCI have an increased risk of developing dementia,[14,15] which has a negative effect on patient well-being, caregiver burden, and health care costs, level of depressive symptoms, and instrumental activities of daily living.[16–21] The frequency of PD-MCI is variable, ranging from 19% to 55%, and is present even in early, de novo, untreated PD patients.[13] Hence, the identification of PD-MCI is clinically important for the management of patient care and prognosis. The MDS Task Force criteria conceptualize two levels of diagnostic methods: an abbreviated assessment (Level I) or a comprehensive assessment (Level II), which can also classify

[1]Charles University in Prague, First Faculty of Medicine and General University Hospital, Department of Neurology and Center of Clinical Neuroscience, Prague, Czech Republic; [2]Charles University in Prague, First Faculty of Medicine and General University Hospital, Department of Psychiatry, Prague, Czech Republic

***Correspondence to:** Dr. Ondrej Bezdicek, Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine and General University Hospital in Prague, Charles University in Prague, Kateřinská 30, 128 21 Praha 2, Czech Republic; E-mail: ondrej.bezdicek@gmail.com
Keywords: mild cognitive impairment, Parkinson's disease, validity.
Relevant disclosures and conflicts of interest are listed at the end of this article.
Received 21 October 2015; revised 27 April 2016; accepted 10 May 2016.
Published online 8 July 2016 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/mdc3.12391

doi:10.1002/mdc3.12391

PD-MCI subtypes as single or multiple domains impaired. However, studies analyzing the classification and diagnostic accuracy of the MDS PD-MCI neuropsychological battery at Level II (MDS neuropsychological battery) and the differences between PD-MCI subtypes are thus far scarce.[22–26]

In accord with the MDS Diagnostic Criteria for PD-MCI at Level II and the framework for assessing cognitive impairment in PD,[8,27] the principal goal of the present study was to evaluate the discriminative validity of the best neuropsychological tests for cognitive impairment in PD-MCI in comparison to patients with PD without cognitive deficit (PD-ND) and healthy controls (HCs) and normative sample (NS). Additionally, we endeavored to explore the relation between PD-MCI Level I and II and the number of false positive and false negative findings with regard to PD-MCI status.

Patients and Methods

The clinical sample consisted in total of 106 patients with PD, matched with 120 HCs according to age and education (Table 1). All subjects were recruited from the Movement Disorders Center, Department of Neurology, First Faculty of Medicine and General University Hospital in Prague (Prague, Czech Republic). All PD patients were examined by a neurologist specialized in movement disorders and met the United Kingdom PD Society Brain Bank criteria.[31] Exclusion criteria were as follows: PDD according to MDS criteria[10]; atypical or secondary parkinsonism; severe or unstable depression; with psychotic manifestations (hallucinations or delusions); anticholinergic medications; and other medical or neurological conditions potentially resulting in cognitive impairment (e.g., seizure, stroke, or head trauma). Levodopa equivalent daily dose (LEDD) for each patient was calculated (Table 1).[32] All PD patients were examined in the ON motor state. The study was approved by the ethics committee of the General University Hospital in Prague, and all participants provided signed, informed consent. All tests were administered under standard neuropsychological laboratory conditions and were conducted by trained psychologists (O.B., T.N.). Participants (patients and paired controls) underwent a comprehensive clinical evaluation that included medical history, evaluation of functional abilities, medication status, motor status by the UPDRS Part III

TABLE 1 Demographic characteristics of HC (n = 120) and PD (n = 106) participants and NS (N = 699)

	HC	(Range)	PD	(Range)	*P* Value
Age	58.03 ± 8.59	(41–82)	59.93 ± 8.41	(41–82)	0.094[a]
Education	14.60 ± 2.87	(9–25)	14.12 ± 2.95	(8–21)	0.223[a]
Sex (% men)	60		65		0.241[b]
Laterality (% right-handed)	89		97		0.016[b]
Language (% Czech)	100	—	100	—	—
Ethnicity (% Caucasians)	100	—	100	—	—
BDI-II	5.33 ± 3.38	(0–12)	10.77 ± 5.98	(0–31)	<0.001[a]
Duration of PD, years	—		12.08 ± 4.54	(4–30)	—
UPDRS-III	—		16.55 ± 9.00	(2–39)	—
H & Y stage	—		2.06 ± 0.58	(1–3)	—
Medication, mg LEDD	—		1193 ± 637	(565–2,828)	—
	NS	No. (n %)			
Age	25–29	9 (1.3)			
	30–34	20 (2.9)			
	35–39	13 (1.9)			
	40–44	26 (3.7)			
	45–49	30 (4.3)			
	50–54	28 (4.0)			
	55–59	31 (4.4)			
	60–64	93 (13.3)			
	65–69	91 (13.0)			
	70–74	95 (13.6)			
	75–79	79 (11.3)			
	80–84	90 (12.9)			
	85–89	66 (9.4)			
	90+	28 (4.0)			
Education	Lower	185 (26.5)			
	Higher	514 (73.5)			
Laterality	Right-handed	636 (91.0)			
	Others	63 (9.0)			
Sex	Female	354 (50.6)			
	Male	345 (49.4)			

Mean and SD with range in parentheses or percentages reported. LEDD and dopamine agonist equivalent daily dose (DAED) for each patient were calculated on the basis of theoretical L-dopa equivalents, as follows: L-dopa dose + L-dopa dose × 1/3 if on entacapone + L-dopa dose × 1/2 if on tolcapone, pramipexole (mg) × 100 + ropinirole (mg) × 20 + rotigotine (mg) × 30 + selegiline (mg) × 10 + rasagiline (mg) × 100 + amantadine × 1 + apomorphine (mg) × 10. NS: other = left handed, ambidextrous, or nondominant right-handed (left-handed from birth but instructed to write with nondominant right hand); lower = level of formal education as measured by number of years spent at school (8–11 years of formal education); higher = college-level or higher-level education (12 and more years).
[a]Independent-samples *t* test.
[b]χ^2 test.

doi:10.1002/mdc3.12391

(UPDRS-III), and standard MDS Level II neuropsychological assessment for the diagnosis of PD-MCI.[8] The neuropsychological battery at Level I (abbreviated assessment) generally consists of global scales of cognitive abilities,[8] in the present study of the Mattis Dementia Rating Scale, Second Edition (DRS-II).[33] Level II (comprehensive assessment) consisted of 10 tests in five cognitive domains; one measure from each test was derived as recommended:[8] (1) attention and working memory (Digit Span [DS] forward from the Wechsler Adult Intelligence Scale, Third Revision [WAIS-III] and Trail Making Test [TMT]-A)[34]; (2) executive function (TMT-B and Prague Stroop test interference condition)[34,35]; (3) language (semantic fluency [animals, clothes, and shopping] and WAIS-III Similarities)[36]; (4) memory (Rey Auditory Verbal Learning Test [RAVLT], learning delayed recall and Family Pictures [FP] delayed recall from the Wechsler Memory Scale, Third Revision [WMS-III])[37]; and (5) visuospatial function (CLOX and Judgment of Line Orientation).

PD-MCI Classification

First, we used the demographically matched (according to age and education) control sample (n = 120; HCs) to divide PD group (n = 106) to PD-MCI and PD-ND at Levels I and II. We normalized the raw scores of the HC group from each of the 10 measures in the battery (that included tests recommended by the MDS PD-MCI task force; further only as "battery") and DRS-II (Level I) and transformed them to *z*-scores using the Rankit formula (rankits of a set of data are the expected values of the order statistics of a sample from the standard normal distribution).[38] On Level II, patients that scored 1.5 standard deviations (SDs) or 2.0 SDs[24] below the average *z*-score derived from the HC group were classified as PD-MCI. These subjects satisfied impairment criteria on at least two neuropsychological tests, represented by either two impaired tests in one cognitive domain or one impaired test in two different cognitive domains.[8] On Level I, patients that scored 1.5 SDs or 2.0 SDs[24] below the average DRS-II *z*-score derived from the HC group were classified as PD-MCI (Table 2). Subsequently, Level I (to prevent circularity) was applied for PD-MCI and PD-ND classification, and this classification was used for the determination of discriminative validity for each battery test measure (Level II). Receiver operating characteristic (ROC) curves (ROC) with area under the ROC curve (AUC) and 95% confidence intervals (CIs) were calculated.

To strengthen our analyses, we adopted the same procedure for the classification of PD-MCI and PD-ND according to Levels I and II that scored 1.5 SD or 2.0 SD[24] below the average *z*-score of the NS (N = 699; Tables 1 and 3). These transformations were based on previous regression analysis of all 10 measures in the battery based on NS. Accordingly, Levels I and II were applied for the classification of PD-MCI and PD-ND, and this classification was used for the determination of discriminative validity based on the ROC analysis for each battery test measure (Level II) and battery composite score (averaged *z*-score from 10 tests; Table 4). Accordingly, PD-MCI subtypes (single vs. multiple domain impairment and further subtypes

TABLE 2 Comparison of PD-MCI and PD-ND classifications according to algorithms based on Level I and II MDS criteria and HC as comparison group (for SD)

Classification Criterion	Group	Level II Battery −2.0 SDs		Level II Battery −1.5 SDs	
		PD-MCI (n = 50)	PD-ND (n = 56)	PD-MCI (n = 64)	PD-ND (n = 42)
Level I DRS-II −2.0 SDs	PD-MCI (n = 23)	21	2	22	1
	PD-ND (n = 83)	29	54	42	41
Level I DRS-II −1.5 SDs	PD-MCI (n = 35)	28	7	32	3
	PD-ND (n = 71)	22	49	32	39

TABLE 3 Comparison of PD-MCI and PD-ND classifications according to algorithms based on Level I and II MDS criteria and normative data (for SD)

Classification Criterion	Group	Level II Battery −2.0 SDs		Level II Battery −1.5 SDs	
		PD-MCI n = 40	PD-ND n = 66	PD-MCI n = 54	PD-ND n = 52
Level I DRS-II −2.0 SDs	PD-MCI n = 29	19	10	21	8
	PD-ND n = 77	21	56	33	44
Level I DRS-II −1.5 SDs	PD-MCI n = 35	22	13	27	8
	PD-ND n = 71	18	53	27	44

TABLE 4 MDS PD-MCI Level II neuropsychological battery discriminative validity in PD-MCI vs. PD-ND based on normative data comparisons (NS; N = 699) in the Czech population

	Below −2.0 SDs *d*	Below −1.5 SDs *d*
1. Attention and working memory		
TMT-A	0.56[a]	0.70[a]
DS backward	0.66[a]	0.64[a]
2. Executive function		
TMT-B	0.83[a]	0.89[a]
PST-C	0.57[a]	0.59[a]
3. Language		
Semantic fluency	0.77[a]	0.66[a]
WAIS-R Sim	1.06[a]	0.81[a]
4. Memory		
RAVLT-DR	0.59[a]	0.67[a]
WMS-III FP-DR	0.42[a]	0.60[a]
5. Visuospatial functions		
JoL	0.85[a]	0.62[a]
CLOX-I	0.26[a]	0.30[a]
Composite score	1.12[a]	1.14[a]

d = Cohen`s effect size coefficient. MDS PD-MCI battery domains, raw scores for the neuropsychological tests are provided. 2. Executive function = Letter fluency (N + K + P).
[a]$P < 0.05$.
PST-C, Prague Stroop Test Interference Score; Semantic fluency (animals+ clothes + shopping); WAIS-R Sim, Similarities; RAVLT-DR, Rey Auditory Verbal Learning Test Delayed Recall; WMS-III FP-DR, Wechsler Memory Scale, Third Revision Family Pictures Delayed Recall; JoL, Judgment of Line Orientation; CLOX-I, Royall's CLOX

doi:10.1002/mdc3.12391

based on attention and working memory, executive function, language, memory, and visuospatial functions) were identified as suggested by MDS PD-MCI criteria.[8,23]

HC participants were recruited from the general community through advertisements (nonrandom sampling), and a brief medical history for each subject was obtained by telephone. A cohort of 120 healthy subjects was included (Table 1), having met the following criteria for enrollment: Interviews excluded all participants with a history of head trauma with loss of consciousness, cerebrovascular accident, abuse of alcohol or other psychoactive substances, and individuals with a history of neurological or psychiatric disease or ongoing delirium. We additionally excluded persons currently undergoing radio- or chemotherapy, with a major medical condition (myocardial infarction, diabetes mellitus, etc.), or sensory deficits. Participants meeting the above criteria were then tested for cognitive performance using the Montreal Cognitive Assessment Czech version (MoCA), for manifestations of depression using the Beck Depression Inventory, Second Edition (BDI-II), and instrumental activities of daily living (IADLs) using the Functional Activities Questionnaire (FAQ).[28] To exclude subjects with cognitive impairment, limits for enrollment were set on the MoCA at ≤1.5 SDs in comparison to Czech normative MoCA data.[29] To exclude subjects with a higher level of depression, the BDI-II score was limited at <13, and with respect to impaired IADLs, the FAQ cutoff was set at ≤4.[28,30]

NS participants were recruited for a separate study from the general community through advertisements, and a brief medical history for each subject was obtained by telephone. A cohort of 699 healthy subjects was included with the same inclusion criteria as HCs. Demographic and other cohort characteristics of the NS are presented in Table 1.

Statistical Analyses

All analyses were performed using IBM software (SPSS 22.0 for Windows; IBM Corp, Armonk, NY). Continuous variables are expressed as mean, SD, and range, categorical variables as percentages, and ordinal variables as medians. Normality was evaluated by visual inspection of Q-Q plots and the Shapiro-Wilk test. The Spearman correlation coefficient was used to evaluate the relationship between different ordinal test measures and the Pearson correlation coefficient for different continuous test measures. When assessing the comparison of different groups, we used effect-size statistics (Cohen's *d*) as proposed by Cohen.[40] ROC curves with AUC and 95% CIs were calculated and used in comparing the diagnostic classification of PD-MCI. Level of significance was set at $\alpha = 0.05$.

Results

Classification table based on cutoffs derived from the control sample (HC) at Levels I and II and 1.5 or 2.0 SDs below average can be found in Table 2 (compare also Table S1). An analogous classification algorithm at Levels I and II based on the NS are depicted in Table 3.

The magnitude of difference based on effect size between PD-MCI and PD-ND (classified according to Level I to prevent circularity) in the battery measures and battery composite score is presented in Table 4. The comparison is based on *z*-scores in each measure derived from the normative data.

Moreover, we used the ROC analysis to support the battery effect-size analysis from Table 4, and we show the discriminative potential of all battery test measures at Level II in Table 5 and 6 (see also Fig. S1). All group comparisons are based on

TABLE 5 Detection potential of all tests at Level II based on DRS-II classification (Level I) and different thresholds in HC vs. PD-MCI and PD-ND vs. PD-MCI in different cognitive domains

	PD-MCI Classification DRS-II Cutoff = –2.0 SDs		PD-MCI Classification DRS-II Cutoff = –1.5 SDs	
	HC (n = 120) vs. PD-MCI (n = 23)	PD-ND (n = 83) vs. PD-MCI (n = 23)	HC (n = 120) vs. PD-MCI (n = 35)	PD-ND (n = 71) vs. PD-MCI (n = 35)
1. Attention and WM				
TMT-A	0.85[a] (0.77–0.94)[b]	0.72 (0.60–0.83)[c]	0.81[a] (0.73–0.89)[b]	0.68 (0.57–0.78)[c]
DS backward	0.85[a] (0.77–0.93)[b]	0.75 (0.65–0.86)[b]	0.84[a] (0.77–0.91)[b]	0.78 (0.69–0.87)[b]
2. Executive function				
TMT-B	0.92[a] (0.86–0.98)[b]	0.80[a] (0.71–0.89)[b]	0.89[a] (0.83–0.95)[b]	0.78 (0.69–0.87)[b]
PST-C	0.76 (0.64–0.87)[b]	0.70 (0.57–0.82)[c]	0.76 (0.67–0.86)[b]	0.74 (0.64–0.84)
3. Language				
Semantic fluency	0.94[a] (0.9–0.99)[b]	0.90[a] (0.84–0.96)[b]	0.88[a] (0.81–0.94)[b]	0.86[a] (0.78–0.93)[b]
WAIS-R Sim	0.82[a] (0.73–0.92)[b]	0.84[a] (0.74–0.93)[b]	0.75 (0.66–0.84)[b]	0.78 (0.69–0.88)[b]
4. Memory				
RAVLT-DR	0.84[a] (0.75–0.93)[b]	0.80[a] (0.70–0.91)[b]	0.74 (0.64–0.84)[b]	0.70 (0.59–0.81)[c]
WMS-III FP-DR	0.82[a] (0.73–0.90)[b]	0.74 (0.63–0.86)[b]	0.75 (0.67–0.84)[b]	0.67 (0.56–0.78)[c]
5. Visuospatial functions				
JoL	0.78 (0.66–0.91)[b]	0.67 (0.52–0.82)[c]	0.79 (0.69–0.89)[b]	0.71 (0.59–0.83)[c]
CLOX-I	0.90[a] (0.80–1.00)[b]	0.55 (0.41–0.69)	0.92[a] (0.85–0.99)[b]	0.53 (0.42–0.65)

2. Executive function = Letter fluency (N + K + P).
[a]Test measures with AUC ≥0.80.
[b]$P < 0.001$; [c]$P < 0.05$.
PST-C, Prague Stroop Test Interference Score; Semantic fluency (animals+ clothes + shopping); WAIS-R Sim, Similarities; RAVLT-DR, Rey Auditory Verbal Learning Test Delayed Recall; WMS-III FP-DR, Wechsler Memory Scale, Third Revision, Family Pictures Delayed Recall; JoL, Judgment of Line Orientation; CLOX-I, Royall's CLOX.

doi:10.1002/mdc3.12391

raw scores. The highest discriminative potential based on AUC was detected in PD-ND versus PD-MCI only in TMT-B, semantic fluency, WAIS-R Similarities, and RAVLT delayed recall when only these measures showed high (AUC ≥0.80) discriminative potential. The same pattern was observed at the lower threshold of 1.5 SDs below average, where in PD-ND versus PD-MCI only semantic fluency reached AUC ≥0.80, in PD-MCI versus HC only TMT-A, DS backward, TMT-B, semantic fluency, and CLOX I (Table 5). The highest discriminative potential based on AUC was detected between PD-MCI and HCs, where most tests reached AUC ≥0.80 with a cutoff of 2.0 SDs below average (TMT-A, DS backward, TMT-B, semantic fluency, WAIS-R Similarities, RAVLT delayed recall, WMS-III FP, and CLOX I). In Table 6, we see that these measures, especially TMT-B and semantic fluency, have also high sensitivity and specificity and positive and predictive values.

Discussion

Our study represents the first step in applying and validating the MDS PD-MCI Level I and II criteria in the Czech population with several significant findings. The principal goal of the study was to evaluate the discriminative validity of the MDS PD-MCI neuropsychological battery: First, based on Cohen's *d* statistics for effect size and a comparison of PD-MCI with normative data, we see that there is a very large difference (above 1.0) for the composite score of the MDS neuropsychological battery. Furthermore, there are large differences (above 0.8 for WAIS-R Similarities and TMT-B) for both –1.5 and –2.0 SD thresholds, with most of the tests reaching medium effect size. These results can further strengthen the case for high discriminative validity of the battery in the differentiation of PD-MCI and the importance of large normative data for group comparisons of predementia states, such as PD-MCI.[26,40]

Second, different detection thresholds had a significant impact on the discriminative validity of test measures used in the battery. When comparing PD-MCI to PD-ND by –2.0 SDs in the battery, only semantic fluency reached 0.90 and AUC ≥0.80 (TMT-B, WAIS-R Similarities, and RAVLT delayed recall). At a cutoff of –1.5 SDs, only semantic fluency reached AUC ≥0.80. A comparison between PD-MCI and HCs revealed that below a cutoff of –2.0 SD, executive, language, and visuospatial functional measures (TMT-B, semantic fluency, and CLOX-I) reached superior discriminative validity (AUC ≥0.90). The TMT-A, DS backward, WAIS-R Similarities, and RAVLT delayed recall, as well as WMS-III FP delayed recall, showed very good detection potential (AUC ≥0.80) in the differentiation of PD-MCI from HCs. Lowering the threshold to –1.5 SDs led to a reduction in detection potential (only CLOX-I reached AUC = 0.90), with only TMT-A, TMT-B, DS backward, and semantic fluency reaching AUC ≥0.80. We would therefore recommend, in accord with previous studies, a threshold of –2.0 SD below the normative standard for differentiating PD-MCI from HCs as well as PD-MCI from PD-ND.[24] Furthermore, we offer optimal diagnostic cutoffs (cutoffs with maximum combined sensitivity and specificity) for all 10 measures in the battery, including sensitivity and specificity and positive and negative predictive values, that can be used in clinical practice or in cross-cultural comparisons.

Third, we can see the overlap between classificatory systems (Table 3) even when the comparison is based on large normative standards. PD-MCI can be determined at Level I or II; however, this leads to differing classifications and has implications for type I error (detecting an effect that is not present, i.e., PD-MCI status when the patient is not PD-MCI) and type II error (failing to detect an effect that is present, i.e., indicating that the patient is PD-ND when he or she is PD-MCI). For example, using Level I and the cutoff of 1.5 SDs below average, 71 subjects were classified as PD-ND (77%) and 35 as PD-MCI (33%), whereas using Level II and 1.5 SDs below average, 52 subjects were classified as PD-ND (49%) and 54 as PD-MCI (51%), which indicates higher sensitivity of Level II compared to Level I. Moreover, we see the number of false-positive PD-MCI at Level I (type I error), when 8 of these 35 classified as PD-MCI at Level I are classified as PD-ND at Level II. By this way, we can directly evaluate the congruence between Level I and II and different thresholds (–1.5 and –2.0 SDs). We found that both Levels I and II had partly noncongruent classification results and were not interchangeable. However, the number of cases is very similar or identical when using the normative sample (for a different view of this question when only a comparison group of 120 HCs is used, see Table 2). To be more precise, Table 3 indicates that Level I is very good in controlling for type I error (the incorrect rejection of a true null hypothesis) in the detection of PD-MCI cases, but not good in controlling type II error (the failure to reject a false null hypothesis); that is, sensitivity is low with a number of false-negative cases. Furthermore, after using normative data, there is a higher degree of congruence between Levels I and II. Recommendations for clinical practice, considering the results presented in Table 3, would be to use Level I first, and if the case falls below a given threshold (1.5 or 2.0 SDs below average) on the DRS-II, there is no need to examine the diagnosis of PD-MCI at Level II. Level II is rather needed for a more detailed investigation of the diagnosis of PD-ND determined at Level I.

In general, regarding PD-MCI subtyping into cognitive domains, when comparing PD-MCI with PD-ND, we see a picture of executive-language impairment (semantic fluency and TMT-B), with PD-MCI subtypes represented mostly by single-domain as well as multiple-domain impairment.[8,23]

We would also like to emphasize that semantic fluency (except for the battery composite score) was the most discriminative test measure in all comparisons (independent of PD-MCI vs. HCs or PD-MCI vs. PD-ND). Thus, one may hypothesize that PD-MCIs are most impaired in executive and language (semantic) abilities; however, cognitive impairment in PD-MCI in general covers multiple domains (abnormalities on at least one test in two or more cognitive domains).[8] In addition to executive (TMT-B) memory (RAVLT delayed recall and WMS-III FP delayed recall), these domains include attention and working memory (TMT-A and DS backward) and visuospatial function

doi:10.1002/mdc3.12391

TABLE 6 Discriminative validity of the MDS-NB measures in the detection of PD-MCI including optimal diagnostics cutoffs, sensitivity, specificity, and positive predictive value and negative predictive value

	PD-MCI Classification > DRS-II Cutoff = –2.0 SDs										PD-MCI Classification > DRS-II Cutoff = –1.5 SDs									
	HC (n = 120) vs. PD-MCI (n = 23)					PD-ND (n = 83) vs. PD-MCI (n = 23)					HC (n = 120) vs. PD-MCI (n = 35)					PD-ND (n = 71) vs. PD-MCI (n = 35)				
	Cutoff	Se	Sp	PPV	NPV	Cutoff	Se	Sp	PPV	NPV	Cutoff	Se	Sp	PPV	NPV	Cutoff	Se	Sp	PPV	NPV
1. Attention and WM																				
TMT-A	42/43	0.87	0.70	0.74	0.84	42/43	0.87	0.52	0.64	0.80	46/47	0.66	0.82	0.79	0.71	39/40	0.83	0.50	0.62	0.75
DS backward	5/6	0.78	0.83	0.82	0.79	6/7	0.96	0.43	0.63	0.91	5/6	0.77	0.83	0.82	0.78	5/6	0.77	0.66	0.69	0.74
2. Executive function																				
TMT-B	116/117	0.87	0.85	0.85	0.87	116/117	0.87	0.65	0.71	0.83	100/101	0.91	0.75	0.78	0.89	103/104	0.89	0.59	0.68	0.84
PST-C	41/42	0.48	0.98	0.96	0.65	40/41	0.48	0.89	0.81	0.63	38/39	0.49	0.93	0.88	0.65	38/39	0.49	0.90	0.83	0.64
3. Language																				
Semantic fluency	51/52	0.87	0.93	0.93	0.88	52/53	0.87	0.87	0.87	0.87	64/65	0.94	0.57	0.69	0.90	57/58	0.77	0.80	0.79	0.78
WAIS-R Sim	18/19	0.61	0.89	0.85	0.70	21/22	0.74	0.77	0.76	0.75	22/23	0.74	0.61	0.65	0.70	22/23	0.74	0.69	0.70	0.73
4. Memory																				
RAVLT-DR	6/7	0.78	0.80	0.80	0.78	6/7	0.78	0.79	0.79	0.78	6/7	0.60	0.80	0.75	0.67	6/7	0.60	0.80	0.75	0.67
WMS-III FP-DR	23/24	0.70	0.83	0.80	0.73	27/28	0.78	0.62	0.67	0.74	26/27	0.60	0.79	0.74	0.66	27/28	0.66	0.63	0.64	0.65
5. Visuospatial functions																				
JoL	23/24	0.59	0.81	0.76	0.66	21/22	0.47	0.79	0.69	0.60	23/24	0.63	0.81	0.77	0.69	22/23	0.59	0.78	0.73	0.66
CLOX I	11/12	0.81	0.98	0.98	0.84	11/12	0.81	0.39	0.57	0.67	12/13	0.84	0.92	0.91	0.85	13/14	0.91	0.26	0.55	0.74

Cutoff = optimal diagnostic cutoffs (cutoffs with maximum combined sensitivity and specificity). 2. Executive function = Letter fluency (N + K + P).
Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; PST-C, Prague Stroop Test Interference Score; Semantic fluency (animals + clothes + shopping); WAIS-R Sim, Similarities; RAVLT-DR, Rey Auditory Verbal Learning Test Delayed Recall; WMS-III FP-DR, Wechsler Memory Scale, Third Revision Family Pictures Delayed Recall; JoL, Judgment of Line Orientation; CLOX-I Royall's CLOX.

doi:10.1002/mdc3.12391

(CLOX-I). In this heterogeneous, multiple-domain concept of cognitive impairment in PD-MCI, the most striking feature is the role of semantic fluency. On functional MRI, semantic fluency activates the temporal lobes and the role of the hippocampus and temporal lobes in connection with semantic fluency has recently been well established.[41] The hippocampus is especially activated during semantic fluency and semantic memory tasks.[41] Accordingly, recent studies regarding the nature of cognitive impairment in PD have emphasized the role of memory deficits observed in nondemented PD patients, interpreted largely as the result of learning deficits associated with cortical thinning in the temporoparietal regions and reduced temporal lobe connectivity.[25,42–46] Our results are in accord with this evidence.

The following limitations of the present study must be addressed. First, the patient sample size could potentially be larger, which limits the generalizability of the results in comparison to meta-analytic efforts for the validation of PD-MCI criteria.[26] Another point is the composition of the neuropsychological battery, where the Boston Naming Test could have been used instead of semantic fluency, and the Brief Visuospatial memory Test–Revised instead of WMS-III FP, which may potentially have better discriminative potential. More broadly, the sheer number of possible neuropsychological tests available is large,[47] and the relevance of diagnosis by individual tests may be limited; however, we tried to compute also the composite *z*-score for all tests at Level II. Another limitation is that our study is cross-sectional, and we are not able to trace the evolution of cognitive impairment in PD. Third, even though our PD-MCI classification is based on our recent and culturally adapted DRS-II normative data and the clinical utility of the DRS-II in PD-MCI and PD-D research was established,[34,48–50] the limitations of the DRS-II and selected individual neuropsychological tests may lie in their overlap in several cognitive domains, for example, semantic fluency is a substantial part of the Conceptualization subscale in the DRS-II. Fourth, for PD-MCI classification, we used DRS-II (i.e., Level I); however, the gold standard of Level II in this article would have to be determined using another gold standard, for instance, the predictive value for PDD in longitudinal perspective. This approach clearly limits the generalizability of the present study regarding Level II and each of the battery measures and their cutoffs.

In conclusion, the present study supports the validity and feasibility of MDS PD-MCI diagnostic criteria. An empirical investigation of the discriminative validity of the PD-MCI battery at Level II showed semantic fluency as the most useful and discriminative measure. We also demonstrated the potential caveats of a Level I and II classification approach and of different cut-off thresholds (–1.5 or –2.0 SDs) with their impact on classification and type and I and II errors based on comparison group and normative standards.

Author Roles

(1) Research Project: A. Conception, B. Organization, C. Execution; (2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique; (3) Manuscript Preparation: A. Writing of the First Draft, B. Review and Critique.

O.B.: 1A, 1B, 1C, 2A, 3A
T.N.: 1C, 2C, 3B
J.M.: 2A, 2B
F.R.: 1C, 2C, 3B
P.H.: 1C, 2C, 3B
J.R.: 1C, 2C, 3B
R.J.: 1C, 2C, 3B
E.R.: 1C, 2C, 3B

Acknowledgments

The authors thank all participants for their willingness to take part in the present study and administrators (Nela Zemanová, MA, Lenka Ošlejšková, MA, Tereza Maková, MA, and Hana Žaloudková, MA) for their help with data collection.

Disclosures

Ethical Compliance Statement: We confirm that we have read the journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

Funding Sources and Conflicts of Interest: This work was supported the by the Czech Ministry of Education grants IGA NT12282-5 and LH13256 (VES13-KontaktII), PRVOUK-P26/LF1/4, and GAUK 920413 and by the Alzheimer Foundation Fund. The authors report no conflicts of interest.

Financial Disclosures for previous 12 months: The authors declare that there are no disclosures to report.

References

1. Jankovic J. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* 2008;79:368–376.
2. Owen AM. Cognitive dysfunction in Parkinson's disease: the role of frontostriatal circuitry. *Neuroscientist* 2004;10:525–537.
3. Williams-Gray CH, Mason SL, Evans JR, et al. The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. *J Neurol Neurosurg Psychiatry* 2013;84:1258–1264.
4. Williams-Gray CH, Evans JR, Goris A, et al. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain* 2009;132(Pt 11):2958–2969.
5. Williams-Gray CH, Foltynie T, Brayne CE, Robbins TW, Barker RA. Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. *Brain* 2007;130(Pt 7):1787–1798.
6. Kehagia AA, Barker RA, Robbins TW. Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with Parkinson's disease. *Lancet Neurol* 2010;9:1200–1213.
7. Hely MA, Reid WG, Adena MA, Halliday GM, Morris JG. The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Mov Disord* 2008;23:837–844.
8. Litvan I, Goldman JG, Troster AI, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord* 2012;27:349–356.
9. Dubois B, Burn D, Goetz C, et al. Diagnostic procedures for Parkinson's disease dementia: recommendations from the Movement Disorder Society Task Force. *Mov Disord* 2007;22:2314–2324.
10. Emre M, Aarsland D, Brown R, et al. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord* 2007;22:1689–1707; quiz, 1837.

doi:10.1002/mdc3.12391

11. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004;256:183–194.

12. Troster AI. A precis of recent advances in the neuropsychology of mild cognitive impairment(s) in Parkinson's disease and a proposal of preliminary research criteria. *J Int Neuropsychol Soc* 2011;17:393–406.

13. Aarsland D, Bronnick K, Williams-Gray C, et al. Mild cognitive impairment in Parkinson disease: a multicenter pooled analysis. *Neurology* 2010;75:1062–1069.

14. Pedersen KF, Larsen JP, Tysnes OB, Alves G. Prognosis of mild cognitive impairment in early Parkinson disease: the Norwegian ParkWest study. *JAMA Neurol* 2013;70:580–586.

15. Broeders M, de Bie RM, Velseboer DC, Speelman JD, Muslimovic D, Schmand B. Evolution of mild cognitive impairment in Parkinson disease. *Neurology* 2013;81:346–352.

16. Schrag A, Jahanshahi M, Quinn N. What contributes to quality of life in patients with Parkinson's disease? *J Neurol Neurosurg Psychiatry* 2000;69:308–312.

17. Aarsland D, Larsen JP, Karlsen K, Lim NG, Tandberg E. Mental symptoms in Parkinson's disease are important contributors to caregiver distress. *Int J Geriatr Psychiatry* 1999;14:866–874.

18. Aarsland D, Bronnick K, Ehrt U, et al. Neuropsychiatric symptoms in patients with Parkinson's disease and dementia: frequency, profile and associated care giver stress. *J Neurol Neurosurg Psychiatry* 2007;78:36–42.

19. Vossius C, Larsen JP, Janvin C, Aarsland D. The economic impact of cognitive impairment in Parkinson's disease. *Mov Disord* 2011;26:1541–1544.

20. Pirogovsky E, Schiehser DM, Obtera KM, et al. Instrumental activities of daily living are impaired in Parkinson's disease patients with mild cognitive impairment. *Neuropsychology* 2014;28:229–237.

21. Yarnall AJ, Breen DP, Duncan GW, et al. Characterizing mild cognitive impairment in incident Parkinson disease: the ICICLE-PD study. *Neurology* 2014;82:308–316.

22. Biundo R, Weis L, Pilleri M, et al. Diagnostic and screening power of neuropsychological testing in detecting mild cognitive impairment in Parkinson's disease. *J Neural Transm (Vienna)* 2013;120(4):627–633.

23. Goldman JG, Weis H, Stebbins G, Bernard B, Goetz CG. Clinical differences among mild cognitive impairment subtypes in Parkinson's disease. *Mov Disord* 2012;27:1129–1136.

24. Goldman JG, Holden S, Bernard B, Ouyang B, Goetz CG, Stebbins GT. Defining optimal cutoff scores for cognitive impairment using Movement Disorder Society Task Force criteria for mild cognitive impairment in Parkinson's disease. *Mov Disord* 2013;28:1972–1979.

25. Stefanova E, Žiropadja L, Stojković T, et al. Mild Cognitive impairment in early Parkinson's disease using the Movement Disorder Society Task Force Criteria: cross-sectional study in Hoehn and Yahr Stage 1. *Dement Geriatr Cogn Disord* 2015;40:199–209.

26. Geurtsen GJ, Hoogland J, Goldman JG, et al. Parkinson's disease mild cognitive impairment: application and validation of the criteria. *J Parkinsons Dis* 2014;4:131–137.

27. Marras C, Troster AI, Kulisevsky J, Stebbins GT. The tools of the trade: a state of the art "How to Assess Cognition" in the patient with Parkinson's disease. *Mov Disord* 2014;29:584–596.

28. Bezdicek O, Stepankova H. Martinec Novakova L, Kopecek M. Toward the processing speed theory of activities of daily living in healthy aging: normative data of the Functional Activities Questionnaire. *Aging Clin Exp Res* 2016;28:239–247.

29. Kopecek M, Stepankova H, Lukavsky J, Ripova D, Nikolai T, Bezdicek O. Montreal Cognitive Assessment (MoCA): normative data for old and very old Czech adults. *Appl Neuropsychol Adult* 2016. doi: 10.1080/23279095.2015.1065261. [Epub ahead of print]

30. Bezdíček O, Lukavský J, Preiss M. Validizační studie české verze dotazníku FAQ [Functional Activities Questionnaire, Czech version: a validation study]. *Cesk Slov Neurol N* 2011;74/107:36–42.

31. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992;55:181–184.

32. Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov Disord* 2010;25:2649–2653.

33. Bezdicek O, Michalec J, Nikolai T, et al. Clinical validity of the Mattis Dementia Rating Scale in differentiating mild cognitive impairment in Parkinson's disease and normative data. *Dement Geriatr Cogn Disord* 2015;39:303–311.

34. Bezdicek O, Moták L, Axelrod BN, et al. Czech version of the Trail Making Test: normative data and clinical utility. *Arch Clin Neuropsychol* 2012;27:906–914.

35. Bezdicek O, Lukavsky J, Stepankova H, et al. The Prague Stroop Test: Normative standards in older Czech adults and discriminative validity for mild cognitive impairment in Parkinson's disease. *J Clin Exp Neuropsychol* 2015;37:794–807.

36. Nikolai T, Štěpánková H, Michalec J, et al. Testy verbální fluence, česká normativní studie pro osoby vyššího véku [Verbal fluency tests. Czech normative study for older persons]. *Cesk Slov Neurol N* 2015;78/111:292–299.

37. Bezdicek O, Stepankova H, Moták L, et al. Czech version of Rey Auditory Verbal Learning test: normative data. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2014;21:693–721.

38. Solomon SR, Sawilowsky SS. Impact of rank-based normalizing transformations on the accuracy of test scores. *J Mod Appl Stat Meth* 2009;8:448–462.

39. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum; 1988.

40. Shirk SD, Mitchell MB, Shaughnessy LW, et al. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimers Res Ther* 2011;3:32.

41. Glikmann-Johnston Y, Oren N, Hendler T, Shapira-Lichter I. Distinct functional connectivity of the hippocampus during semantic and phonemic fluency. *Neuropsychologia* 2015;69:39–49.

42. Chiaravalloti ND, Ibarretxe-Bilbao N, DeLuca J, et al. The source of the memory impairment in Parkinson's disease: acquisition versus retrieval. *Mov Disord* 2014;29:765–771.

43. Tanner JJ, Mareci TH, Okun MS, Bowers D, Libon DJ, Price CC. Temporal lobe and frontal-subcortical dissociations in non-demented Parkinson's disease with verbal memory impairment. *PLoS ONE* 2015;10:e0133792.

44. Bronnick K, Alves G, Aarsland D, Tysnes OB, Larsen JP. Verbal memory in drug-naive, newly diagnosed Parkinson's disease. The retrieval deficit hypothesis revisited. *Neuropsychology* 2011;25:114–124.

45. Mak E, Su L, Williams GB, et al. Baseline and longitudinal grey matter changes in newly diagnosed Parkinson's disease: ICICLE-PD study. *Brain* 2015;138(Pt 10):2974–2986.

46. Pirogovsky-Turk E, Filoteo JV, Litvan I, Harrington DL. Structural MRI correlates of episodic memory processes in Parkinson's disease without mild cognitive impairment. *J Parkinsons Dis* 2015;5:971–981.

47. Goldman JG, Holden S, Ouyang B, Bernard B, Goetz CG, Stebbins GT. Diagnosing PD-MCI by MDS Task Force criteria: how many and which neuropsychological tests? *Mov Disord* 2015;30:402–406.

48. Villeneuve S, Rodrigues-Brazete J, Joncas S, Postuma RB, Latreille V, Gagnon JF. Validity of the Mattis Dementia Rating Scale to detect mild cognitive impairment in Parkinson's disease and REM sleep behavior disorder. *Dement Geriatr Cogn Disord* 2011;31:210–217.

49. Kulisevsky J, Pagonabarraga J. Cognitive impairment in Parkinson's disease: tools for diagnosis and assessment. *Mov Disord* 2009;24:1103–1110.

50. Llebaria G, Pagonabarraga J, Kulisevsky J, et al. Cut-off score of the Mattis Dementia Rating Scale for screening dementia in Parkinson's disease. *Mov Disord* 2008;23:1546–1550.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Neuropsychological (Level II) characteristics of PD (n = 106) and HC (n = 120) participants and thresholds for cognitive impairment including effect size

Fig. S1. Detection potential of four most discriminative tests at Level II based on ROC analysis in PD-ND (n = 83) versus PD-MCI (n = 23) at –2 SDs.

doi:10.1002/mdc3.12391