Retrospective Assessment of Movement Disorder Society Criteria for Mild Cognitive Impairment in Parkinson's Disease

Andrea M. Loftus,[1] Romola S. Bucks,[1,2] Meghan Thomas,[3,4] Robert Kane,[1] Caitlin Timms,[1] Roger A. Barker,[5] AND Natalie Gasson[1]

[1]ParkC, Curtin Neuroscience Laboratory, School of Psychology and Speech Pathology, Curtin University, Perth, Western Australia
[2]School of Psychology, University of Western Australia, Perth, Western Australia
[3]Medical Sciences, Edith Cowan University, Perth, Western Australia
[4]Experimental and Regenerative Neuroscience, School of Animal Biology, University of Western Australia, Western Australia
[5]Department of Clinical Neuroscience, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom

Abstract

A Movement Disorder Society (MDS) taskforce recently proposed diagnostic criteria for Parkinson's disease with features of mild cognitive impairment (PD-MCI). This study first examined the prevalence and nature of PD-MCI in a non-demented cohort using the MDS criteria. Using the generic Monte Carlo simulation method developed by Crawford and colleagues (2007), this study then estimated the base rate of the representative population who would demonstrate PD-MCI due to chance alone. A total of 104 participants with idiopathic PD underwent extensive motor and neuropsychological testing at baseline and 2 years later. The Unified Parkinson's Disease Rating Scale (UPDRS) was used to assess motor symptoms of PD and a range of established neuropsychological tests was used to assess PD-MCI in accord with MDS criteria. In accord with MDS criteria, 38% of this cohort demonstrated PD-MCI at baseline and 48% at follow-up. Of the 36 participants in the multiple-domain PD-MCI subtype at time-1, 9 (25%) demonstrated no PD-MCI at follow up. Analysis revealed that approximately 13% of the representative population would demonstrate abnormally low scores for 2 of the 9 tests used, thereby meeting MDS criteria for PD-MCI. Clinicians and researchers need to approach a single diagnosis (i.e., based on one assessment) of PD-MCI with considerable caution. (*JINS*, 2015, *21*, 137–145)

Keywords: Parkinson's disease, Cognition, Mild cognitive impairment, Movement disorders, Neuropsychological tests, Dementia

INTRODUCTION

Mild cognitive impairment is now acknowledged as a feature of Parkinson's disease (PD-MCI). PD-MCI has a reported prevalence of 25–30% in PD without dementia and is associated with increased age, disease duration, and severity (Aarsland et al., 2010). PD-MCI increases the risk of deve loping PD dementia (PD-D) relative to no PD_MCI, although the rate and pattern of cognitive decline remains unclear (Aarsland, Tandberg, Larsen, & Cummings, 1996; Janvin, Larsen, Aarsland, & Hugdahl, 2006; Williams-Gray, Foltynie, Brayne, Robbins, & Barker, 2009). PD-MCI negatively affects functional independence (Rosenthal et al., 2010) and quality of life (Klepac, Trkulja, Relja, & Babić, 2008;

Schrag, Jahanshahi, & Quinn, 2000). In light of its impact and potential to predict PD-D, the early identification of PD-MCI is a priority.

Janvin and colleagues (2006) assessed PD-MCI using three tests (Benton Visual Retention Test, BVRT; Stroop; Judgment of Line Orientation, JLO). An abnormal score was defined as 1.5 standard deviations (*SD*) below the mean of the normative sample. Janvin et al. identified three MCI subtypes; (i) amnestic MCI (impaired BVRT), (ii) multiple-domain MCI (≥2 impaired tests), and (iii) single non-memory MCI (impaired on Stroop *or* JLO). Of the 72 non-demented participants, 53% demonstrated PD-MCI: most having single, non-memory impairment (one deficit, non-memory test; 44.7%). Four years later, 62% of those with PD-MCI had progressed to dementia, compared to 6% of those without, leading the authors to suggest that PD-MCI represents the initial stage of cognitive decline in dementia. However, no information was provided regarding

how many of those no longer met criteria for PD-MCI: that is, were later identified as cognitively normal.

Aarsland et al. (2010) examined cognitive impairment in eight cohorts, testing across three cognitive domains, including memory (visual and verbal), executive/attention, and visuospatial functioning. As this was a multicenter study, the number and type of tasks differed, ranging from a minimum 7 to a maximum 11 tasks. The executive/attention cognitive domain typically comprised the highest number of tests. Using regression (and resulting intercept and regression weights) based on domain Z scores in the control group, the authors calculated expected cognitive domain scores for participants. Participants were identified as having PD-MCI if the difference between their predicted and actual score was more than 1.5 SDs below the mean for at least 1 of the 3 domains. Of the 1346 participants included in the study, 347 (25.8%) were diagnosed with PD-MCI, which varied across centers from 18.9–39.4%. Memory was the most commonly impaired domain (13.3%), followed by visuospatial (11%) and executive function (10.1%). The most prevalent PD-MCI type involved impairment in only one cognitive domain (20% of participants), followed by two-domain impairment (5%) and three-domain impairment (1%). Although memory was the most commonly impaired single-domain, a greater number of participants demonstrated non-amnestic single-domain impairment than amnestic. However, memory contained the least tests, which may have impacted the findings.

Approximately a quarter of people with non-demented PD demonstrate PD-MCI (26.7%; Litvan et al., 2011). Litvan et al. (2011) reported that impairments occurred across a range of cognitive domains and that non-amnestic was more common than amnestic PD-MCI. Litvan et al. also suggested that deficits in cognitive function can be present even at the time of diagnosis, indicating that those in the early stages of disease development are also at risk of PD-MCI. Litvan and colleagues suggested that PD-MCI is common, heterogeneous in nature and is a risk factor for progression to PD-D.

There has been a lack of standardized, diagnostic criteria for PD-MCI. In response, a Movement Disorder Society (MDS) taskforce recently proposed formal, diagnostic criteria for PD-MCI (see Litvan et al., 2012). The cognitive criteria include two levels of assessment, which vary in terms of their scope and efficacy. Level 1 comprises a shortened assessment using a global measure or a limited number of neuropsychological tests (just one per cognitive domain), the findings of which do not support a comprehensive diagnosis of PD-MCI. Level 2 requires at least two tests for each of the five cognitive domains of (i) executive function, (ii) attention and working memory, (iii) language, (iv) memory, and (v) visuospatial function. The level 2 criteria allow for subtyping of PD-MCI as either single or multiple-domain. Two abnormal tests within one cognitive domain and unimpaired tests in the other four indicates single-domain impairment. At least one abnormal test in two or more domains indicates multiple-domain impairment. Identification of the impaired domain is also recommended, so that potential differences between sub-types may be explored.

The MDS level 2 criteria recommend the use of at least two neuropsychological tests for each of the five cognitive domains, suggesting the use of *at least ten tests* for the assessment of PD-MCI. If a participant demonstrates impairment on 2 of the minimum 10 tests, they would meet the criteria for PD-MCI. According to MDS criteria, an abnormal test score would be 1–2 standard deviations (*SD*) below the appropriate mean. Defining an abnormally low score as one which falls, for example, 1.5 *SD* below the norm, suggests that approximately 5% of the normative population will demonstrate an abnormal score (Type I error). This raises a critical issue in neuropsychological testing – the use of assessment comprising multiple tests and the associated increased risk of Type I errors.

Neuropsychological assessments comprising multiple tests should consider the number of people from the representative population who would be predicted to demonstrate an abnormally low test score (Crawford, Garthwaite, & Gault, 2007). For assessments using multiple tests, a higher number of the normal population would be expected to demonstrate an abnormally low test score. In light of this, what is typically defined as an abnormally low test score may not actually be that unusual in the normal population. Indeed, substantial numbers of a normal population can demonstrate abnormally low test score(s) using the percentile/*SD* criteria definition of impairment (Crawford et al., 2007) when multiple tests are combined.

Where multiple tests are used, Crawford and colleagues (2007) recommend that researchers should first estimate the percentage of the normal population who would be expected to demonstrate abnormally low test scores. By providing information about the rate of Type I error associated with the cognitive assessment, this estimate assists in the interpretation of a participant's performance and helps to avoid over-inference about the presence of abnormal test scores (Crawford et al., 2007). Crawford and colleagues propose a Monte Carlo simulation method to estimate the base rate of a population who would demonstrate abnormally low tests (see Crawford et al., 2007, for a detailed description). Using this method, researchers can estimate the percentage of participants who would be expected to demonstrate abnormally low test scores for a given number of tests.

The present study retrospectively applied the Level 2 MDS taskforce criteria (excepting language, for which there was only one test) to examine PD-MCI subtypes in a non-demented, community-based cohort over a 2-year period. The aims of the study were several-fold. First, to investigate the prevalence and type of PD-MCI in accordance with the MDS taskforce criteria. Second, to examine the stability of PD-MCI subtypes over time by retrospectively applying the MDS criteria at baseline and 2 years later. Finally, to use Crawford's method to estimate the percentage of the population who would be expected to demonstrate 2 abnormally low tests scores (i.e., meet the MDS criteria for PD-MCI).

METHODS

Participants

Participants were recruited as part of an ongoing study at ParkC in Western Australia. This study was approved by a University ethics committee and all research was conducted in accordance with the Declaration of Helsinki. All participants provided written, informed consent.

Participants were included in the study only if they had a formal diagnosis of Idiopathic PD by a neurologist or geriatrician, in accordance with the United Kingdom Parkinson's Disease Society Brain Bank Clinical (UKPDSBBC). Participants were excluded from the study if they self-reported any condition which may interfere with cognitive assessment.

As this study was a retrospective application of the MDS criteria for PD-MCI, the specific MDS inclusion-exclusion criteria were applied wherever possible. Participants were included if they or an informant had reported problems with their cognition. For most participants (72), this was quantified using section I of the Unified Parkinson's Disease Rating Scale-Revised (UPDRS-Rev; Goetz et al., 2008): The Non-Motor Aspects of Experiences of Daily Living Scale (nM-EDL). Of those participants for whom the nM-EDL was available, only those who reported that their cognitive experiences were not sufficient to interfere with their functional independence were included in the present study. Of those participants for whom nM-EDL was not available at time-1 (32), participants were included on the basis that they (i) attended and completed the testing session (indicative of some degree of functional independence) and (ii) demonstrated cognitive deficits on formal neuro-psychological testing.

One hundred four participants with idiopathic PD were included. In accordance with Emre et al. (2007), no participants had PD-D at study entry. The mean Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975) score at study entry was 28.5 (*SD* = 1.2), range 24–30, suggesting no global cognitive impairment was present.

General Procedure

Participants were mailed a questionnaire pack before undertaking a neuropsychological and motor assessment at ParkC. All participants were tested in the "on" state - approximately 1 hr post-PD-medications. Assessments took ~2.5 hr. Participants completed a baseline assessment (time-1) and a follow-up assessment 2 years later (time-2).

Measures

Age, sex, employment status, marital status, level of education, age at disease onset, disease duration, and medication use were collected *via* self-report questionnaire. Most participants (72 at time-1) completed the UPDRS-Rev at time-1 and all (104) participants completed the UPDRS at time-2. All participants completed the neuropsychological assessment at times 1 and 2.

UPDRS- Section III measured the motor features of PD. The 18 items of section III (motor) align with 6 motor signs of PD (tremor, rigidity, bradykinesia, facial expression, speech, axial impairments). Participants were rated on a scale of 0–4 for 26 different movements. The sum of all items was used: higher scores indicate greater disease severity. Summed scores can range from 0 (asymptomatic) to 104 (most severe).

Mini Mental State Examination (MMSE; Folstein et al., 1975) measured global cognitive functioning. Total MMSE score was calculated using the serial 7s technique (Strauss, Sherman, and Spreen, 2006). Participants were included in the study if they met or exceeded the (age/education appropriate) cutoff scores provided by Iverson (1998) using the norms provided by Crum, Anthony, Bassett, and Folstein (1993).

Australian version of the National Adult Reading Test (AUSNART; Hennessy & McKenzie, 1995) estimated pre-morbid intelligence using a regression equation (Sullivan, Senior, & Hennessy, 2000). The AUSNART was used only to determine the appropriate normative values for some measures.

In accordance with MDS taskforce recommendations, two measures (excepting language, where only one measure was available) were selected for each of the five cognitive domains.

1. Attention and Working Memory

MMSE Serial 7s measured attention and working memory. The participant counts backward by 7s, with each correct subtraction being awarded 1 point. Serial 7s has previously been used as a measure of attention in PD (Janvin et al., 2006) as it does not require a motor or timed response. Normative data were derived from Strauss et al. (2006).

Spatial Working Memory was assessed using the Cambridge Neuropsychological Test Automated Battery (CANTAB[TM]). Participants must find tokens hidden within an array of boxes. Task complexity was determined by the number of boxes present on a trial, which ranged from 3 to 8. Total errors (the number of times a box was selected that was certain not to contain a token) was used. Age/gender appropriate normative data were provided by CANTAB[TM] (2013)

2. Executive

Stockings of Cambridge, a measure of strategy use, was assessed using the CANTAB[TM]. Three colored balls were presented at the top of the screen in a specific configuration. At the bottom of the screen were three identical balls, in a different configuration, which participants were asked to match with the goal set, in the minimum number of moves. Task complexity was determined by the minimum number of moves to solve the task, which ranged from 1 to 5. The proportion of problems solved in the minimum number of moves was calculated across 12 trials (two two-move trials, two three-move, four four-move, and four five-move). Higher

scores indicate greater planning ability. Age-appropriate normative data were provided by CANTAB[TM] 2013.

Controlled Oral Word Association Task (F, A, S; Benton, 1968) measured verbal fluency/generativity. Participants generated as many words as possible beginning with the letters F, A, and S, each in 60 s. Verbal fluency was calculated as the total number of correct responses across all three trials, with higher scores indicating greater verbal fluency. Normative data were derived from Tombaugh, Kozak, and Rees (1999).

3. Language

*Controlled Oral Word Association Task Animals (*Benton, 1968) assessed semantic word knowledge and fluency. Participants named as many animals as they could within 60-s. Participants could name any animal (mammals, reptiles, birds, fish, insects) starting with any letter of the alphabet. The number of correct responses was used, with higher scores indicating greater verbal fluency. Normative data were derived from Tombaugh et al. (1999).

4. Memory

Hopkins Verbal Learning Test – Revised (HVLT-R; Brandt & Benedict, 2001) measured verbal learning and memory. Twelve nouns (four words from three semantic categories) were read to the participant three times. After each reading, the participant was asked to recall as many words as possible. The total correct responses over trials 1–3 was used. Normative data were derived from Hester, Kinsella, Ong, and Turner (2004).

Pattern Recognition Memory was measured using the CANTAB[TM] to assess recognition memory for nonverbal information. Twelve patterns were presented one at a time in the centre of the screen for 3 s each. Participants were instructed to remember them. Forced-choice recognition memory (one target paired with a distractor) was tested after the presentation of all 12 stimuli. The percentage of correct responses was used. Age-appropriate normative data were provided by CANTAB[TM] 2013.

5. Visuospatial

Cube Analysis Sub-test from the Visual Object and Space Perception battery (VOSP; Warrington & James, 1991) measured visuospatial function. The stimuli were black outline representations of a three-dimensional arrangement of square bricks. Each card had a different number of square bricks, constructed in various designs. Participants reported how many square bricks were present. Participants completed two practice trials followed by 10 test trials that were graded in difficulty. Total correct trials were used.

Number Location Sub-test from the VOSP (Warrington & James, 1991) measured visuospatial function. Two boxes were presented on each card. The upper box contained numbers in various locations. The lower box contained one dot in the same location as one of the numbers in the upper box. Participants identified the number that corresponded with the location of the dot in the lower box. Participants completed two practice trials followed by 10 test trials. Total correct trials were used. Normative data for both VOSP tests were derived from Herrera-Guzman, Pena-Casanova, Lara, Gudayol-Ferre, and Bohm (2004).

RESULTS

Applying MDS criteria, an individual test score was deemed abnormal if it was 1.5 *SD* or more below the age/gender/education appropriate normative mean (except error values, where 1.5 *SD* or more above the normative error was deemed abnormal). Participants who demonstrated no deficit test scores or one deficit test score in only one cognitive domain were classified as "No PD-MCI". Participants who demonstrated two abnormal test scores within one cognitive domain and unimpaired test scores in the other four were classified as "single-domain" PD-MCI. Participants who demonstrated at least one abnormal test score in two or more domains were classified as "multiple-domain" PD-MCI. Participants were categorized at time-1 and time-2. Of the 104 participants, 38% demonstrated PD-MCI at time-1 and 48% at time-2. Frequency data, demographics and UPDRS for each subtype at both times are presented in Tables 1 and 2.

Table 1. Frequencies and demographics (mean, standard deviation) for each PD-MCI subtype (as defined by the MDS) at time-1

N = 104	No PD-MCI	Single-domain	Multiple-domain
N	65	3	36
Males (% of subtype)	58.50	66.70	77.80
Age at participation	62.34 (8.93)	70.63 (6.03)	68.53 (8.40)
Age at diagnosis	57.34 (9.82)	62.33 (8.73)	62.31 (10.63)
Disease duration (years)	4.89 (1.03)	8.03 (3.26)	5.74 (1.01)
UPDRS Section III	25.79 (11.45)	34 (11.14)	32.84 (15.22)
Total years of education	14.83 (2.70)	7 (0)	11.57 (2.64)
Retired (% of subtype)	69.20	100	90.10
Married (% of subtype)	87.70	0	75.10

Table 2. Frequencies and demographics (mean, standard deviation) for each PD-MCI subtype (as defined by the MDS) at time-2

N = 104	No PD-MCI	Single-domain	Multiple-domain
N	54	6	44
Males (% of subtype)	66.70	50	65.90
Age at participation	65.89 (8.93)	59 (9.46)	68.72 (9.26)
Age at diagnosis	59.56 (8.52)	51.33 (7.23)	59.84 (12.16)
Disease duration (years)	6.33 (3.70)	7.67 (3.38)	8.89 (5.62)
UPDRS Section III	30.75 (14.51)	26.50 (18.69)	39.92 (18.78)
Total years of education	12.76 (3.88)	12.20 (2.86)	11.95 (3.79)
Retired (% of subtype)	74.10	66.70	95.20
Married (% of subtype)	85.20	83.30	76.70

An independent samples t-test examined whether there were differences between the No-PD-MCI group and the PD-MCI groups combined (an *n* of 3 for the single-domain subtype meant a between-groups analysis could not be conducted). There were no significant demographic differences between the groups (all $ps > .05$)

The off-diagonal elements in Table 3 represent the proportions of participants who migrated from a particular time-1 PD-MCI subtype to a particular time-2 subtype; the diagonal elements represent the participants who maintained their time-1 subtypes at time-2. Those with No PD-MCI at time-1 tended to remain in this subtype at time-2. When they did migrate, it was to the multiple-domain subtype (25%), with fewer migrating to the single-domain subtype (6%).

All three participants in the single-domain subtype at time-1 migrated to the multiple-domain subtype at time-2. These three participants demonstrated two deficit tests in attention/SWM at time-1 and 2, with two participants demonstrating an additional language test deficit at time-2 and another demonstrating an additional visuospatial deficit at time-2. These three participants therefore acquired deficits in another domain beyond the affected time-1 domain (attention/SWM).

Those in the multiple-domain subtype at time-1 tended to remain there at time-2. Two participants migrated to single-domain, both demonstrating one deficit test score in executive and one in attention at time-1. At time-2, both of these participants had no deficit test in executive but had two deficit test scores in attention.

Of the 36 participants in the multiple-domain subtype at time-1, 9 (25% of the subtype, 9% of the total sample; 7 male;

mean age = 64.78; *SD* = 6.47) migrated to No-PD-MCI at time-2; that is they appeared to recover from their PD-MCI. Eight participants demonstrated one abnormal test score in 2 or more domains at time-1 (typically attention/SWM, executive, visuospatial). One participant demonstrated two abnormal test scores in one domain (memory) and one abnormal test score in two other domains (executive, attention) at time-1, but at time-2 demonstrated only one deficit test score in one domain (attention/SWM).

Crawford and colleagues' (2007) generic Monte Carlo simulation method was used to estimate the base rate of the population who would, in accord with MDS criteria, exhibit *j* or more abnormally low tests (see Crawford et al., 2007, for a detailed description). The first step in the simulation is to define the criteria for an abnormally low score and transform this into a standard normal deviate (in the present study, 1.5 *SD*). Following this, the *R* values for the correlations between all the tests used must be entered into the simulation program. These values are used to obtain the Choleski decomposition of *R* ($\sqrt{R}$). The program then generates a random vector of the number of tests used, which is post-multiplied by the Choleski decomposition matrix. These steps are repeated to represent the scores of approximately 1 million cases (in Crawford's program, this process generates 1 million vectors). The number of abnormal scores obtained on each Monte Carlo trial (for each simulated member of the population) is noted. Finally, the program produces a total estimate of those exhibiting *j* abnormal scores across the course of the simulation. This simulation process was conducted five times in the present study; once with the inter-test correlations from our sample and

Table 3. Percentages (numbers) of participants who migrated from a particular time-1 subtype to a particular time-2 subtype (off-diagonal elements) or maintained their time-1 subtype at time-2 (diagonal elements)

Time 1		Time 2	
	No PD-MCI	Single-domain	Multiple-domain
N = 104			
No-PD-MCI (N = 65)	69.23 (45)	6.15 (4)	24.62 (16)
Single-domain (N = 3)	0 (0)	0 (0)	100 (3)
Multiple-domain (N = 36)	25 (9)	5.56 (2)	69.4 (25)

Table 4. Correlations for neuropsychological tests from the ParkC cohort study (357 participants with PD, 41 controls)

Domain	Attention/SWM		Memory		Executive		Visuospatial	
	MMSE Serial 7s	SWM	PRM	HVLT	SOC	Word Fluency	Cube analysis	Number location
Attention/SWM								
MMSE Serial 7s	1							
SWM	−.13*	1						
Memory								
PRM	.27**	−.29**	1					
HVLT	.29**	−.40**	.50**	1				
Executive								
SOC	.30**	−.43**	.39**	.43**	1			
Word Fluency	0	−.29**	−.04	.28**	.15**	1		
Visuospatial								
Cube Analysis	.16**	−.27**	.35**	.32**	.40**	.19**	1	
Number Location	.04	−.16**	.08	.19**	.22**	.07	.32**	1
Language								
Category Fluency	.02	−.30**	.17**	.38**	.28**	.60**	.29**	.15**

once with the correlations set to zero, and again with all correlations set to 0.3, 0.5 and 0.7 so as to demonstrate the robustness of the findings to sample-based variations in inter-test correlations. Data from 357 people with PD and 41 healthy controls (398 participants in total; ParkC cohort) were used as the normative sample to generate the between-test correlations from which base rates are estimated, and estimates were calculated using a computer program by Crawford et al. (2007). The preferred criterion for an abnormally low score was 1.5 SD below the mean for the entire sample. The correlations between the 9 tests used in the present study are presented in Table 4.

Correlations between tests were used to determine the number of participants in the population who would be expected to demonstrate abnormally low test scores. Applying the MDS criteria of at least two deficit tests (across domains or within the same domain), 13.01% of the sample would demonstrate PD-MCI by chance alone (see Table 5). This suggests that nearly 14 participants would demonstrate two abnormally low test scores, thereby meeting PD-MCI criteria, but potentially would be misclassified. Table 5 provides estimates for the current ParkC cohort for a given number of tests. The estimates for fixed and mean

correlations (0, 0.3, 0.5 and 0.7) are included for comparison. As can be seen, mean correlations of 0 (no correlation) and 0.7 (high correlations) between tests would still result in 10–14% of participants exhibiting 2 abnormally low tests scores by chance alone.

Motor Symptom Severity

Motor severity increased from time-1 to time-2. A paired samples t-test revealed a significant difference between time-1 and time-2 UPDRS scores, $t(72) = -3.26$; $p < .05$, $d = -.36$, reflecting the increased severity of motor features from time-1 ($M = 28.38$; $SD = 13.21$) to time-2 ($M = 33.71$; $SD = 17.33$). Paired-samples t-tests of UPDRS-III scores revealed no significant differences between time-1-time-2 for those who remained in the No PD-MCI subtype, $p > .05$. For those who did not have PD-MCI at time-1 but demonstrated PD-MCI at time-2, there was a significant time-1-time-2 difference in their UPDRS-III scores, $t(17) = -2.78$; $p < .05$, $d = -.66$, reflecting increased motor severity from time-1 ($M = 26.78$; $SD = 13.24$) to time-2 ($M = 36.33$; $SD = 15.45$). This is consistent with the suggestion that PD-MCI is associated with the progression of motor features (among other factors).

Table 5. Percentage of population expected to exhibit j or more abnormally low scores (<1.5SD, or 6.6[th] percentile) as a function of the observed (ParkC cohort) or fixed average correlation between tests

	Average r	Percentage exhibiting j or more abnormally low scores for given number of tests								
		1	2	3	4	5	6	7	8	9
ParkC cohort	.25	41.39	13.01	4.10	1.23	0.32	0.07	0	0	0
Fixed	0	46.31	11.79	1.84	0.19	0.01	0	0	0	0
	.30	35.72	14.45	6.04	2.51	1.02	0.37	0.12	0.03	0.01
	.50	29.01	14.19	7.80	4.42	2.47	1.34	0.66	0.28	0.08
	.70	22.19	12.99	8.65	5.99	4.17	2.85	1.85	1.08	0.49

DISCUSSION

This study examined the prevalence and subtyping of PD-MCI in accordance with the MDS taskforce criteria. Prevalence rates of PD-MCI in the present study were higher than those reported in studies preceding the MDS criteria. There are a couple of potential factors to consider when accounting for this higher rate. First, as the present study used a greater number and range of neuropsychological tests than those used by previous studies of PD-MCI, it may have been more sensitive to cognitive deficits. Alternatively, by requiring 10+ tests, the MDS criteria may be associated with increased Type I error rates. The higher prevalence rate reported in the present study could, therefore, be a natural consequence of the total number of tests used and the correlations between those tests.

The present study estimated the percentage of a representative population who would be expected to demonstrate 2 abnormally low tests scores. Approximately 13% of the present sample would be expected to meet MDS criteria for PD-MCI due to Type I error. This number is consistent with the percentage of participants in the present study who, despite demonstrating multiple-domain PD-MCI at time-1, demonstrated no PD-MCI at time-2. Unstable diagnosis of MCI is not unique to PD. Koepsell, Gill, and Chen (2013) reported that MCI diagnoses were the least stable in their multi-centre, longitudinal study from 32 Alzheimer's Disease Centers (Koepsell et al., 2013).

Studies preceding the MDS criteria indicate that single-domain PD-MCI is most common (Aarsland et al., 2010; Rosenthal et al., 2010), with attention (Foltynie, Brayne, Robbins, & Barker, 2004) and memory (Aarsland et al., 2010) being primarily impacted. In the present study, only 3 participants demonstrated single-domain PD-MCI: all three with non-amnestic deficits in attention. In Aarsland's study, non-amnestic single-domain PD-MCI may have been more common than amnestic single-domain because there were more non-amnestic measures used, whereas we used two measures per domain (excepting language). It may be, therefore, that Aarsland et al. found more non-amnestic single-domain cases for methodological reasons, rather than because this is reflective of MCI phenotypes in PD.

Consistent with Broeders et al. (2013), multiple-domain was the most common PD-MCI subtype in this study. As discussed previously, this may be due either to the nature of multiple-domain involvement or the inflation of multiple-domain diagnoses by virtue of the number of tests. The most impacted cognitive domain was attention/spatial working memory, which is consistent with previous research and is thought to reflect frontal decline (Foltynie et al, 2004; Williams-Gray at al., 2007). Executive deficits were also evident, which may be associated with dopaminergic dysfunction in frontostriatal networks (Broeders et al., 2013; Foltynie et al., 2004).

Two factors may explain the movement between multiple-domain and No-PD-MCI observed in the present study. First, the multiple-domain subtype may be less stable than the single-domain subtype. A relatively high number of participants moved from multiple-domain at time-1 to No-PD-MCI at time-2. In comparison, all of the single-domain participants at time-1 moved to multiple-domain at time-2; having acquired additional deficit(s) in another domain. None of the single-domain participants was later identified as No-PD-MCI. Second, those who moved from multiple-domain to No-PD-MCI may be the 13% of participants who would be expected to show two abnormal test scores as a consequence of the number of tests used. If this is the case, these participants were incorrectly diagnosed with multiple-domain PD-MCI at time-1. Incorrect diagnosis of even 1 in 10 individuals with MCI is problematic, having implications for what people with PD will be told about their likely future prognosis, what irreversible steps they may take in relation to employment, pension or family finances, and the potential to cause unnecessary distress. Further application of the MDS criteria to other PD cohorts is required to examine the stability of the multiple-domain subtype in PD-MCI.

Also worthy of consideration is the impact of missing data in longitudinal studies of PD-MCI. The progression to PD-D can occur rapidly and may result in missing follow-up data for some participants. Missing data may have impacted previous results such that single-domain PD-MCI was more likely to be detected than multiple-domain, especially if data were missing across a couple of select domains. Missing data were not an issue in the present study, as all participants completed all cognitive tests. All cognitive domains, excepting language for which there was just the one measure, were represented equally. Of note, none of the participants demonstrated any deficit on this task, despite the fact that category fluency is highly sensitive to the breakdown of language structures in other conditions, such as Alzheimer's disease (Verma & Howard, 2012). This suggests we have not missed any participants with a language-driven presentation. Hence, we suggest that the inclusion of only one test of language is not a substantive issue for the present study. Nonetheless, prospective studies established to apply the MDS Taskforce criteria will be able to confirm or refute this. Of note, however, had we been able retrospectively to use 10 tests as recommended, the chance of two scores being in the abnormal range would have increased.

By virtue of being retrospective, the present study was limited in its consideration of contributing factors to PD-MCI. For example, future studies may wish to explore the impact of PD medication regimens, although the impact of dopaminergic medications on cognitive performance is unclear. Perhaps, more importantly, the severity of motor features (as indicated by UPDRS scores and Hoehn and Yahr staging) was a significant contributory factor to PD-MCI in the study by Aarsland et al. (2010). Those without PD-MCI in the study by Aarsland et al. demonstrated less severe motor symptoms compared to those with PD-MCI. This was also the case in the present study, whereby those without PD-MCI demonstrated significantly lower motor severity than those with PD-MCI (regardless of subtype). This suggests that, although information about dopaminergic medication is

highly desirable, motor severity is a key contributor to PD-MCI, beyond that of medication, or is a covariate.

This is the first longitudinal application of the MDS criteria for PD-MCI. Findings suggest that, at least for some individuals, the label PD-MCI may be premature. Whereas these criteria have had the benefit of standardizing the way that the assessment of diagnosis of MCI in PD is approached, and we applaud this initiative since it makes comparison of studies more meaningful, the results reported here suggest that clinicians and researchers need to approach a single diagnosis (i.e., based on one assessment) of PD-MCI with caution. There are several suggestions which are worthy of future discussion. The present study used a 1.5 or greater standard deviation (SD) from the appropriate norm for an abnormal test score, which is the criterion most commonly used in PD-MCI studies (Aarsland et al., 2010; Broeders et al., 2013; Janvin et al., 2006; Leroi, Pantula, McDonald, & Harbishettar, 2012; Poletti et al., 2012). A more conservative abnormal test score criterion, such as 2 SDs from the normative mean, may be associated with different prevalence rates and may reduce the risk of Type I errors. Alternatively, if the MDS criteria were amended to require *at least* two abnormal test scores in two domains, this would reduce the type I error rate to between 2 and 6%. This is a much more acceptable error rate and worthy of future discussion. This reduction, however, only holds for the identification of multiple-domain PD-MCI; the error rate associated with single-domain PD-MCI would be unchanged. The criteria could be further amended to require three abnormal tests within a single-domain for the identification of single-domain PD-MCI, but this strategy increases the burden of assessment for the participant. Continued follow-up of the present cohort will inform about the pattern of cognitive decline preceding PD-D and may contribute toward the development of a predictive model of PD-MCI progression to PD-D.

ACKNOWLEDGMENTS

REFERENCES

Aarsland, D., Bronnick, K., Williams-Gray, C., Weintraub, D., Marder, K., Kulisevsky, J., … Allcock, L. (2010). Mild cognitive impairment in Parkinson disease: A multicenter pooled analysis. *Neurology*, 75, 1062–1069. doi:10.1007/s11910-011-0203-1

Aarsland, D., Tandberg, E., Larsen, J.P., & Cummings, J.L. (1996). Frequency of dementia in Parkinson's disease. *Archives of Neurology*, 53, 538–542.

Benton, A.L. (1968). Differential behavioural effects of frontal lobe disease. *Neuropsychologia*, 6, 53–60.

Brandt, J., & Benedict, R.H.B. (2001). *Hopkins verbal learning test revised*. Lutz, FL: Psychological Assessment Resources.

Broeders, M., de Bie, R.M., Velseboer, D.C., Speelman, J.D., Muslimovic, D., & Schmand, B. (2013). Evolution of mild cognitive impairment in Parkinson Disease. *Neurology*, 81(4), 346–352. doi:10.1212/WNL.0b013e31829c5c86.

Crawford, J.R., Garthwaite, P.H., & Gault, C.B. (2007). Estimating the percentage of thepopulation with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology*, 21(4), 419–430. doi:10.1037/0894-4105.21.4.419.

Crum, R.M., Anthony, J.C., Bassett, S.S., & Folstein, M.F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *Journal of the American Geriatric Society*, 269, 2386–2391.

Emre, M., Aarsland, M.D., Brown, R., Burn, D., Duyckaerts, C., Mizuno, Y., … Dubois, B. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders*, 22(12), 1689–1707.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

Foltynie, T., Brayne, C.E., Robbins, T.W., & Barker, R.A. (2004). The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. *Brain*, 127, 550–560. doi:10.1093/brain/awh067

Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., … Dodel, R. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, 23, 2129–2170. doi:10.1002/mds.22340

Hennessy, M., & MacKenzie, B. (1995). AUSNART: The development of an Australian version of the NART. In J. Fourez & N. Page (Eds.), *Treatment issues and long term outcomes: Proceedings of the Annual Brain Impairment Conference* (pp. 183–188). Bowen Hill: Australian Academic Press.

Herrera-Guzman, I., Pena-Casanova, J., Lara, J.P., Gudayol-Ferre, E., & Bohm, P. (2004). Influence of age, sex and education on the visual object and space perception battery (VOSP) in a healthy normal elderly population. *The Clinical Neuropsychologist*, 18, 385–394.

Hester, R.L., Kinsella, G.J., Ong, B., & Turner, M. (2004). Hopkins verbal learning test: Normative data for older Australian adults. *Australian Psychologist*, 39, 251–255. doi:10.1080/00050060 412331295063

Iverson, G.L. (1998). Interpretation of Mini-Mental State Examination scores in community-dwelling elderly and geriatric patients. *International Journal of Geriatric Psychiatry*, 13, 661–666.

Janvin, C.C., Larsen, J.P., Aarsland, D., & Hugdahl, K. (2006). Subtypes of mild cognitive impairment in Parkinson's disease: Progression to dementia. *Movement Disorders*, 21, 1343–1349. doi:10.1177/0891988705277540

Koepsell, T.D., Gill, D.P., & Chen, B. (2013). Stability of clinical etiologic diagnosis in dementia and mild cognitive impairment: Results from a multicenter longitudinal database. *American Journal of Alzheimer's Disease and Other Dementias*, 28(8), 750–758.

Klepac, N., Trkulja, V., Relja, M., & Babić, T. (2008). Is quality of life in non-demented Parkinson's disease patients related to cognitive performance? A clinic-based cross-sectional study. *European Journal of Neurology*, *15*, 128–133. doi:10.1111/j.1468-1331.2007.02011.x

Leroi, I., Pantula, H., McDonald, K., & Harbishettar, V. (2012). Neuropsychiatric symptoms in Parkinson's disease with mild cognitive impairment and dementia. *Parkinson's Disease*, *2012*, 1–10.

Litvan, I., Aarsland, D., Adler, C.H., Goldman, J.G., Kulisevsky, J., Mollenhauer, B., … Weintraub, D. (2011). MDS task force on mild cognitive impairment in Parkinson's disease: Critical review of PD-MCI. *Movement Disorders*, *26*, 1814–1824. doi:10.1002/mds.23823

Litvan, I., Goldman, J.G., Tröster, A.I., Schmand, B.A., Weintraub, D., Petersen, R.C., … Williams-Gray, C.H. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Movement Disorders*, *27*, 349–356. doi:10.1002/mds.24893

Poletti, M., Frosini, D., Pagni, C., Baldacci, F., Nicoletti, V., Tognoni, G., … Bonuccelli, U. (2012). Mild cognitive impairment and cognitive-motor relationships in newly diagnosed drug-naive patients with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *83*, 601–606.

Rosenthal, E., Brennan, L., Xie, S., Hurtig, H., Milber, J., Weintraub, D., … Siderowf, A. (2010). Association between cognition and function in patients with Parkinson disease with and without dementia. *Movement Disorders*, *25*, 1170–1176. doi:10.1002/mds.23073

Schrag, A., Jahanshahi, M., & Quinn, N. (2000). What contributes to quality of life in Parkinson's Disease? *Journal of Neurology, Neurosurgery, and Psychiatry*, *69*, 308–312.

Strauss, E., Sherman, E.M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). Oxford: Oxford University Press.

Sullivan, R., Senior, G., & Hennessy, M. (2000). *Australian age-education and premorbid cognitive intellectual estimates for the WAIS-III.* Paper presented at the 6th National Conference of the APS College of Clinical Neuropsychologists, Hunter Valley, New South Wales, Australia.

Tombaugh, T.N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal learning: FAS and animal naming. *Archives of Clinical Psychology*, *14*, 167–177.

Verma, M., & Howard, R.J. (2012). Semantic memory and language dysfunction in early Alzheimer's disease: A review. *International Journal of Geriatric Psychiatry*, *27*(12), 1209–1217.

Warrington, E.K., & James, M. (1991). *The visual object and space perception battery*. Bury St Edmunds, United Kingdom: Thames Valley Test Company.

Williams-Gray, C.H., Foltynie, T., Brayne, C., Robbins, T., & Barker, R. (2007). Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. *Brain*, *130*, 1787–1798. doi:10.1093/brain/awm111