Predicting Progression to Parkinson's Disease Dementia Using Multivariate Normative Comparisons

Agelink van Rentergem, J.A.; de Vent, N.R.; Huizenga, H.M.; Murre, J.M.J.; Schmand, B.A.

DOI
10.1017/S1355617719000298

Publication date
2019

Document Version
Final published version

Published in
Journal of the International Neuropsychological Society

Journal of the International Neuropsychological Society (2019), **25**, 678–687

doi:10.1017/S1355617719000298

Predicting Progression to Parkinson's Disease Dementia Using Multivariate Normative Comparisons

J.A. Agelink van Rentergem[1,†], N.R. de Vent[1,†], H.M. Huizenga[1,2,3], J.M.J. Murre[1], ANDI consortium[4] and B.A. Schmand[1,5]

[1]Department of Psychology, University of Amsterdam, Amsterdam, Netherlands
[2]Amsterdam Brain and Cognition Center, University of Amsterdam, Amsterdam, Netherlands
[3]Research Priority Area Yield, University of Amsterdam, Amsterdam, Netherlands
[4]See full list of consortium members on www.andi.nl/
[5]Department of Medical Psychology, Academic Medical Center, Amsterdam, Netherlands

(Received February 1, 2018; Final revision February 18, 2019; Accepted February 18, 2019; First Published Online May 14, 2019)

Abstract

Objective: Parkinson's disease with mild cognitive impairment (PD-MCI) is a risk factor for progression to PD dementia (PDD) at a later stage of the disease. The consensus criteria of PD-MCI use a traditional test-by-test normative comparison. The aim of this study was to investigate whether a new multivariate statistical method provides a more sensitive tool for predicting dementia status at 3- and 5-year follow-ups. This method allows a formal evaluation of a patient's profile of test scores given a large aggregated database with regression-based norms. **Method:** The cognitive test results of 123 newly diagnosed PD patients from a previously published longitudinal study were analyzed with three different methods. First, the PD-MCI criteria were applied in the traditional way. Second, the PD-MCI criteria were applied using the large aggregated normative database. Last, multivariate normative comparisons (MNCs) were made using the same aggregated normative database. The outcome variable was progression to dementia within 3 and 5 years. **Results:** The MNC was characterized by higher sensitivity and higher specificity in predicting progression to PDD at follow-up than the two PD-MCI criteria methods, although the difference in classification accuracy did not reach statistical significance. **Conclusion:** We conclude that MNCs could allow for a more accurate prediction of PDD than the traditional PD-MCI criteria, because there are encouraging trends in both increased sensitivity and increased specificity.

Keywords: Mild cognitive impairment, neuropsychological diagnostics, normative data, regression-based norms, multivariate normative comparison, aggregated database, ANDI

INTRODUCTION

Many Parkinson's disease (PD) patients show a decline in cognitive functioning, often already early in the disease course (Aarsland et al., 2001; Hobson & Meara, 2004; Muslimovic et al., 2005). PD with mild cognitive impairment (PD-MCI) is predictive of progression to PD dementia (PDD; Aarsland et al., 2001; Caviness et al., 2007; Hoogland et al., 2017; Williams-Gray et al., 2007). It is important to accurately predict which patients will develop PDD as it may have implications for patient care, for example, choice of medication (such as avoiding anticholinergic drugs) and planning of assistance. Also, accurate prediction enables a more appropriate selection of patients for cognitive interventions or pharmaceutical trials.

Clinical criteria for PD-MCI have been proposed by a task force of the International Parkinson and Movement Disorder Society (MDS) (Litvan et al., 2012). In order to diagnose PD-MCI at Level II (i.e. the level with most diagnostic certainty), a PD patient should experience subjective complaints (or their relatives should report such complaints) and should be impaired on objective cognitive testing. Litvan et al. (2012) recommend administering at least two tests for each of five cognitive domains, thus a minimum of 10 tests, of which at least two tests need to indicate impairment for a PD-MCI diagnosis. Impairment is usually assessed by comparing the patient's test scores to those of normative samples, often in the form of norm tables that accompany published test manuals.

There are several issues with comparing patients to such published norm tables. First, as the normative data for neuropsychological tests have been collected for each test separately, correlations between tests are usually unknown (except in case of co-normed tests). Because the correlations are unknown, they cannot formally be taken into account in neuropsychological assessment. This makes it hard to evaluate abnormal combinations of scores (i.e. an abnormal score profile; Huizenga et al., 2007). Third, norm tables do not always allow

Correspondence and reprint requests to: N.R. de Vent, Postbus 15916, 1001 NK Amsterdam, The Netherlands. E-mail: n.r.devent@uva.nl
†Denotes shared first authorship

for correction for the influence of demographic variables, even though age, sex, and level of education are known to influence the scores on neuropsychological tests. Moreover, it is often not possible to simultaneously correct for age, sex, and level of education (Lezak et al., 2012). Also, when correction for age is possible, separate norms are presented for different age groups. When a patient gets older and shifts from one age group to the next, the interpretation of their test results can be different and may, for example, change from abnormal to normal (Zachary & Gorsuch, 1985). Fourth, when evaluating more than one test (at least 10 in the case of Level II PD-MCI diagnosis), the likelihood of obtaining an abnormal score by chance alone increases with the number of tests that have been administered (Binder et al., 2009).

In this study, we apply a new statistical method that circumvents these problems, to detect cognitive abnormality in newly diagnosed PD patients and to predict PDD at later follow-up. This method uses an aggregated normative database of neuropsychological tests (de Vent et al., 2016). Because the database contains data of co-normed neuropsychological tests, correlations between tests can be taken into account. This allows for a so-called multivariate normative comparison (MNC), which evaluates a patient's entire profile of test scores. MNC can detect abnormal combinations of high and low scores in a score profile, which are easily overlooked in a traditional, univariate normative comparison (Crawford & Garthwaite, 2002; Huizenga et al., 2007; Su et al., 2015). The database contains information about demographic variables and thus allows simultaneous correction for age, sex, and level of education. By using regression-based demographic corrections, drastic changes in the interpretation of test scores, when moving from one to another norm table, are prevented. The new statistical method keeps the false positive rate under control, because it entails a single statistical comparison.

In this article, we will discuss (1) using an aggregate normative database instead of norm tables and (2) using it to make MNCs. To examine whether these approaches are a good alternative to traditional (univariate) normative comparisons, we compare their performance to that of the PD-MCI criteria with traditional norms. We use existing data from a longitudinal study conducted by our group (Broeders et al., 2013). First, we compare the ability of the traditional PD-MCI criteria to predict PDD after 3 and 5 years to the same PD-MCI criteria when applied with a large normative database of co-normed tests. Second, we compared the traditional PD-MCI criteria to the new MNC when applied with the same large normative database. Finally, we explored whether the new approach can give insight into which cognitive domains in particular are impaired in PD-MCI patients who progress to PDD compared to those who do not.

METHOD

PD Patients

Participants were 123 patients with newly diagnosed idiopathic PD according to the Gelb criteria who at baseline were younger than 85 years, were not demented, had no history of stroke, and had a score of at least 24 on the Mini-Mental State Examination (MMSE; Folstein et al., 1975). The diagnosis of idiopathic PD was checked at study inclusion and was reevaluated on two occasions by a neurologist specialized in movement disorders (Broeders et al., 2013; Muslimovic et al., 2005). Patients whose diagnosis was revised were not included in the group of 123 patients. After 3 years, the clinical status was missing for 26 patients. After 5 years, information was no longer available for another 24 patients (see Table 1). The institutional review boards of the participating hospitals approved the original study and all patients gave written informed consent. The institutional review boards of the participating hospitals approved the original study by Muslimovic et al. (2005) and Broeders et al. (2013) in accordance with the Helsinki Declaration.

Table 1. Sample characteristics of the PD group from Broeders et al. (2013)

		N	% Men	Age range at baseline
Baseline		123	54%	32–84
	Attrition = 26			
3-year follow-up		97	54%	35–84
	Attrition = 24			
5-year follow-up		73	55%	35–84

PD with mild cognitive impairment

Broeders et al. (2013) applied the PD-MCI Level II criteria (Litvan et al., 2012) as follows: (1) Patient has a PD diagnosis; (2) there is gradual cognitive decline as reported by the patient or observed by the caregiver or the clinician; (3) there is a cognitive deficit on neuropsychological testing; and (4) cognitive deficits do not significantly interfere with functional independence. With respect to the first criterion, all patients in the sample were newly diagnosed PD patients; the diagnosis was checked by the study neurologists at follow-up. With respect to the second criterion, gradual cognitive decline reported by the patient was assessed by two questions, asking whether the patient experienced memory problems or concentration problems. If participants answered either question with "yes" or "sometimes", this was recorded as experiencing subjective complaints. With respect to the third criterion, a score of 1.5 SD below the demographically corrected mean on at least two tests was considered a cognitive deficit. To compensate for the fact that gradual cognitive decline as observed by the caregiver or the clinician from Criterion 2 was not available in our data, patients could also be diagnosed with PD-MCI if they reported no subjective complaints but had impairments (of at least 1.5 SD) on four or more tests. The reasoning behind this choice was that such broad impairments would be noticed by caregivers (Broeders et al., 2013). This choice does mean that more patients can be

Table 2. Characteristics of the neuropsychological test variables in ANDI

	N	% Men	Age range	Demographic variables[a]
Memory				
RAVLT – total (Rey, 1964)	5017	50	18–97	A + S + E
RAVLT – delayed recall (Rey, 1964)	4540	49	18–97	A + S + E
RBMT – Story subtest – immediate recall (Wilson et al., 1983).	346	40	19–90	A + S + E
RBMT – Story subtest – delayed recall (Wilson et al., 1983).	353	40	19–89	A + S+ E
Language				
BNT – 30 item (Kaplan et al., 1983)	467	42	18–89	A + S + E
WAIS-III Similarities (Wechsler, 1997).	274	36	18–80	E
Executive functions				
COWAT (Benton & Hamsher, 1983).	2894	48	18–97	A + S + E
TOL – total movement score (Culbertson & Zillmer, 1998)	62	53	40–80	A
Visuospatial/constructive skills				
JOLO (Benton et al. (1983).	69	54	40–80	S + E
Clock Drawing Test (Royall et al. (1998).	167	46	40–82	E
Attention				
WAIS-R Digit Symbol Test (Wechsler, 1981).	2122	43	18–91	A + S + E
TMT – part A (Reitan, 1992)	3216	47	18–97	A + S + E

RAVLT = Rey Auditory Verbal Learning Test, RBMT = Rivermead Behavioral Memory Test, BNT = Boston Naming Test, WAIS-III = Wechsler Adult Intelligence Scale 3rd edition, COWAT = Controlled Oral Word Association Test, TOL= Tower of London, JOLO = Judgment of Line Orientation, WAIS-R = Wechsler Adult Intelligence Scale Revised edition, TMT = Trail Making Test.

[a]As explained elsewhere (de Vent et al., 2016), an Akaike Information Criteria (AIC) selection procedure was used to estimate which of the three demographic variables to include in regression-based demographic corrections. In this column, A, S, and E indicate whether age, sex, and level of education were included for each variable.

classified as having PD-MCI, including patients without subjective cognitive complaints. Because this choice was nonstandard for the PD-MCI criteria (Litvan et al., 2012), we reran the analyses *post hoc* for the Advanced Neuropsychological Diagnostics Infrastructure (ANDI) data with the stricter original rule, excluding patients without subjective cognitive complaints from the PD-MCI classification (see Supplement 3). With respect to the fourth criterion, patients were excluded if they had a score <24 on the MMSE (Folstein et al., 1975).

PD Dementia

PDD was used as the outcome variable. PDD at 3- and 5-year follow-ups was diagnosed by MDS criteria (Emre et al., 2007). Criteria were as follows: (1) a diagnosis of PD prior to the onset of dementia, (2) MMSE score lower than 24, (3) no depression as measured by the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983), (4) cognitive deficits that are severe enough to interfere with daily functioning, as measured by the Behavioral Assessment of Daily Living (Collin et al., 1988), Schwab and England Scale (Schwab & England, 1969), and Functional Independence Measure (Van Putten et al., 1999), and (5) an abnormal score on at least two of the following tests: clock drawing (Lezak et al., 2012), pentagon copying or serial 7s of the MMSE (Folstein et al., 1975).

Materials

PD patients were tested on five cognitive domains: memory, language, executive functions, visuospatial skills, and attention. All test variables from the Broeders et al. (2013) study were used except the Modified Wisconsin Card Sorting Test (Nelson, 1976) as its score distribution was extremely skewed, violating the assumptions of the parametric normative comparisons that are used throughout this article. We replaced it by the Tower of London (Culbertson & Zillmer, 1998) as an alternative test for the executive functions domain. An overview of the tests can be found in Table 2.

Normative Control Sample

For normative comparisons, we used either the published norms of each neuropsychological test or the database of the ANDI (de Vent et al., 2016). ANDI is an online tool that can be used by clinicians and researchers to conduct normative comparisons. ANDI has a large aggregated normative database ($N = 26{,}635$) which consists of healthy individuals who either participated as control subjects in clinical studies, or participated in community-based studies. Since each participant completed only a subset of the tests that are included in ANDI, the number of participants per test varies between 62 and 5017 depending on the test. Table 2 gives an overview of the tests used for the present analyses and of the number of participants per test. All scores in the ANDI database have been transformed to normality and standardized to

Table 3. Demographic and clinical characteristics for the three groups (PD-MCI criteria, ANDI PD-MCI criteria, and ANDI MNC) at baseline

	PD-MCI criteria		ANDI PD-MCI criteria		ANDI-MNC	
	Normal cognition	PD-MCI	Normal cognition	PD-MCI	Normal cognition	Abnormal cognition
	N = 80(65%)	*N* = 43(35%)	*N* = 90 (73%)	*N* = 33 (27%)	*N* = 91 (74%)	*N* = 32 (26%)
Age	65.1 (10.6)	68.0 (10.1)	64.4 (10.7)	70.8 (8.3)	64.8 (10.1)	69.8 (10.7)
Sex M/F	43/37	23/20	46/44	20/13	48/43	18/14
MMSE	28.0 (1.9)	27.0 (2.0)	28.1 (1.9)	26.5 (1.9)	28.0 (1.8)	26.5 (2.1)
Disease duration in months	18.3 (8.9)	20.1 (13.4)	18.0 (8.8)	21.3 (14.6)	18.2 (8.8)	20.8 (14.9)
LED	139.0 (142.6)	149.9 (139.3)	139.1 (143.5)	152.9 (135.4)	138.4 (145.2)	155.5 (129.7)
UPDRS-III	15.8 (7.8)	19.4 (7.8)	16.2 (8.2)	19.4 (7.0)	16.1 (7.8)	19.9 (7.8)
H&Y	1.6 (0.7)	2.1 (0.7)	1.7 (0.7)	2.1 (0.7)	1.7 (0.7)	2.1 (0.7)
HADS	8.5 (6.6)	13.5 (7.5)	9.4 (7.1)	12.7 (7.6)	9.4 (7.2)	12.8 (7.1)
SE-ADL	91.2 (5.8)	88.1 (7.9)	90.6 (6.9)	89.1 (6.3)	90.2 (7.0)	90.0 (6.2)
BADL	19.7 (0.7)	19.4 (1.5)	19.6 (1.2)	19.6 (0.8)	19.6 (1.2)	19.7 (0.7)

MMSE = Mini-Mental State Examination; LED = LEVODOPA EQUIVALENT DOSE; UPDRS = Unified Parkinson's Disease Rating Scale; H&Y = Hoehn & Yahr scale; HADS = Hospital Anxiety and Depression Scale; SE-ADL = Schwab & England Activities of Daily Living; BADL = Behavioral Assessment of Daily Living.

demographically corrected z scores. For most test variables, age, sex, and level of education had a significant effect, and thus were included in the demographic correction (de Vent et al., 2016).

PD-MCI Criteria Applied with ANDI's Normative Data

In applying the PD-MCI criteria, Broeders et al. (2013) followed typical neuropsychological practice and used normative data from test manuals and various other sources to judge whether a patient deviated from the norm. Here, we applied the PD-MCI Level II criteria in the same way but now with the ANDI database instead of the normative data accompanying each test. A difference with the usual way of working is that the ANDI data have been treated in a consistent manner across all tests (de Vent et al., 2016): For each test, the same procedures were followed for outlier removal, test score standardization, and selection of transformations to normality. Another difference is that for many tests a larger normative sample is available. Instead of *z*-values, ANDI uses *t*-values because these have better statistical properties (Crawford & Garthwaite, 2002). As threshold for abnormality of test scores, we used the clinical convention of −1.5 standard deviations below the demographically corrected mean, which corresponds to a *p* value of .067 one-tailed. Because tests were one-tailed, only deviations in the negative direction were classified as impaired.

Abnormality as Defined by MNC

Finally, MNCs were used (Huizenga et al., 2007) to assess abnormal profiles. MNC compares the profile of the patient's scores to the norm, that is, to the profile of scores that is predicted for a healthy participant of the same age, sex, and level of education (Agelink van Rentergem et al., 2017a, 2017b). The MNC procedure results in a *p* value, which indicates abnormality when it is below a certain threshold. We tested for impairment (one-tailed), that is, only deviations in the negative direction were classified as impaired. In univariate comparisons, if no subjective complaints were present, we required four instead of two significant results. In the MNC, this adaptation is not possible, as only one statistical test result is obtained. Therefore, we used different threshold values for those with and without subjective complaints, to determine whether results were significant. For patients without subjective complaints, we used the same threshold of $p < .067$ (corresponding to −1.5 SD). For patients with subjective complaints, a less strict criterion was chosen, mimicking the procedure for the univariate comparisons. Therefore, we used a *p* value of 2 *.067 one-tailed for this group.

Analysis

We calculated whether the classification of cognitive impairment at baseline is predictive of progression to PDD. Sensitivity and specificity were compared across these three methods. Sensitivity was calculated by dividing the number of patients who were classified as impaired at baseline and progressed to PDD by the total number of patients who developed PDD. Specificity was calculated by dividing the number of patients who were classified as "not-impaired" at baseline and did not progress to PDD by the total number of patients who did not develop PDD. This was done separately for the progression to PDD after 3 years, and after 5 years.

RESULTS

Demographic Characteristics

In Table 3, demographic and clinical characteristics of the patients are given, where the patients are separated into cognitively normal and abnormal categories using each of the three methods.

Table 4. Sensitivity and specificity for progression to PDD of each method (original PD-MCI criteria, PD-MCI applied with ANDI, and MNC method applied with ANDI), specified for 3- and 5-year follow-ups. In parentheses: 90% confidence interval (Agresti & Coull, 1998)

	3-year follow-up		5-year follow-up	
	Sensitivity	Specificity	Sensitivity	Specificity
PD-MCI criteria	.78 (.50–.93)	.68 (.60–.76)	.59 (.39–.76)	.75 (.64–.83)
PD-MCI criteria ANDI	.67 (.40–.86)	.77 (.69–.84)	.47 (.29–.66)	.86 (.76–.92)
MNC ANDI	.89 (.61–.99)	.82 (.74–.88)	.71 (.50–.85)	.89 (.80–.95)

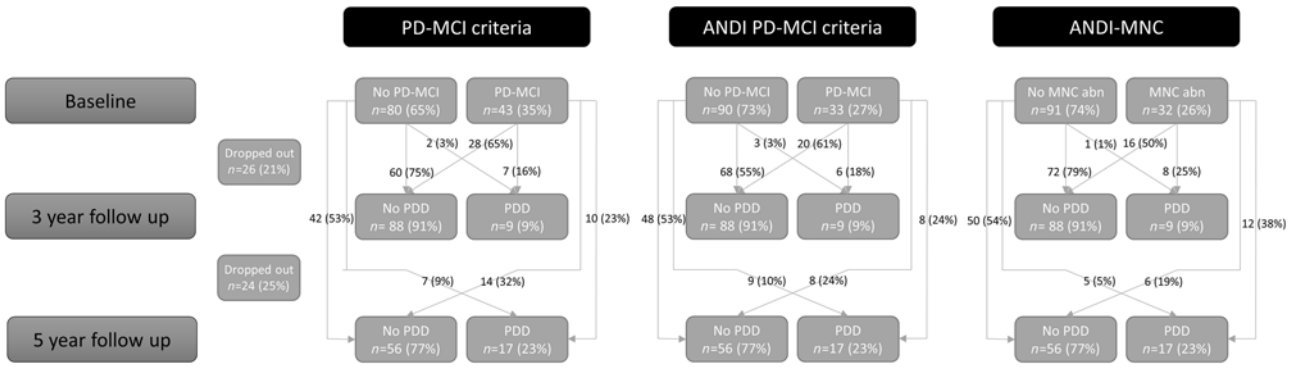

Fig. 1. Progression of PD patients ($n = 123$) to PDD after 3 ($n = 97$) and 5 years ($n = 73$) for the three methods: PD-MCI criteria, PD-MCI criteria applied with ANDI, and MNCs applied with ANDI.

Progression to PDD

Figure 1 gives an overview of progression to PDD for each method evaluated. According to the criteria used by Broeders et al. (2013), at baseline, 35% of the PD patients had PD-MCI. After 3 years, 16% of the PD-MCI patients had progressed to PDD and 65% had not. Of the group who did not have PD-MCI, 3% of patients nevertheless had progressed to PDD and 75% had not (the remaining patients were lost to follow-up). After 5 years, 23% of those with PD-MCI at baseline had progressed to PDD while 32% had not. Of the group who did not have PD-MCI, 9% had progressed to PDD while 53% had not.

The PD-MCI criteria applied with ANDI show that at baseline 27% of the patients had PD-MCI. After 3 years, 18% of the PD-MCI patients had progressed to PDD and 61% had not. Of the group who did not have PD-MCI, 3% had progressed to PDD and 55% had not. After 5 years, 24% of the PD-MCI patients had progressed to PDD and 24% had not. Of the group who did not have PD-MCI, 10% patients had progressed to PDD while 53% had not.

The MNC method applied with the ANDI normative data shows that at baseline 26% PD patients were considered to be MNC abnormal versus 74% who were not. After 3 years, 25% of the MNC abnormal PD patients had progressed to PDD and 50% had not. Of the group who were not MNC abnormal, 1% had progressed to PDD and 79% had not. After 5 years, 38% of the MNC abnormal PD patients had progressed to PDD and 19% had not. Of the group who were not MNC abnormal, 5% patients nevertheless had progressed to PDD while 54% had not.

In Figure 1, it is not visible how much overlap there is in the three different types of diagnostic methods at baseline. For example, the 32 classified as MNC abnormal could theoretically be different patients from those 33 classified as having PD-MCI using the ANDI method. The overlap in diagnoses between pairs of classification methods is explored in Supplement 2. Each of the three methods did indeed differ somewhat in the patients they classified as impaired, although the percentages of agreement were high (78–87%) and kappa's ranged from .49 to .68.

Sensitivity and Specificity

Sensitivity and specificity of the three methods are given in Table 4. After 5 years, sensitivity decreased for all methods but specificity remained similar or increased. However, the confidence intervals are large due to the small size of the samples that remained after attrition. Although differences between the sensitivities and specificities for the three methods did not reach significance (all $ps > .099$), there is a trend showing higher specificity of the MNC method.

Comparing the results of the original PD-MCI criteria to those obtained with the same criteria using ANDI shows a trade-off: sensitivity is higher for the original PD-MCI criteria, and specificity is higher for the PD-MCI criteria applied with ANDI. Therefore, there does not seem to be a clear advantage for either method.

MNCs as applied with ANDI fare better than the other two methods. At both the 3-year and the 5-year follow-ups,

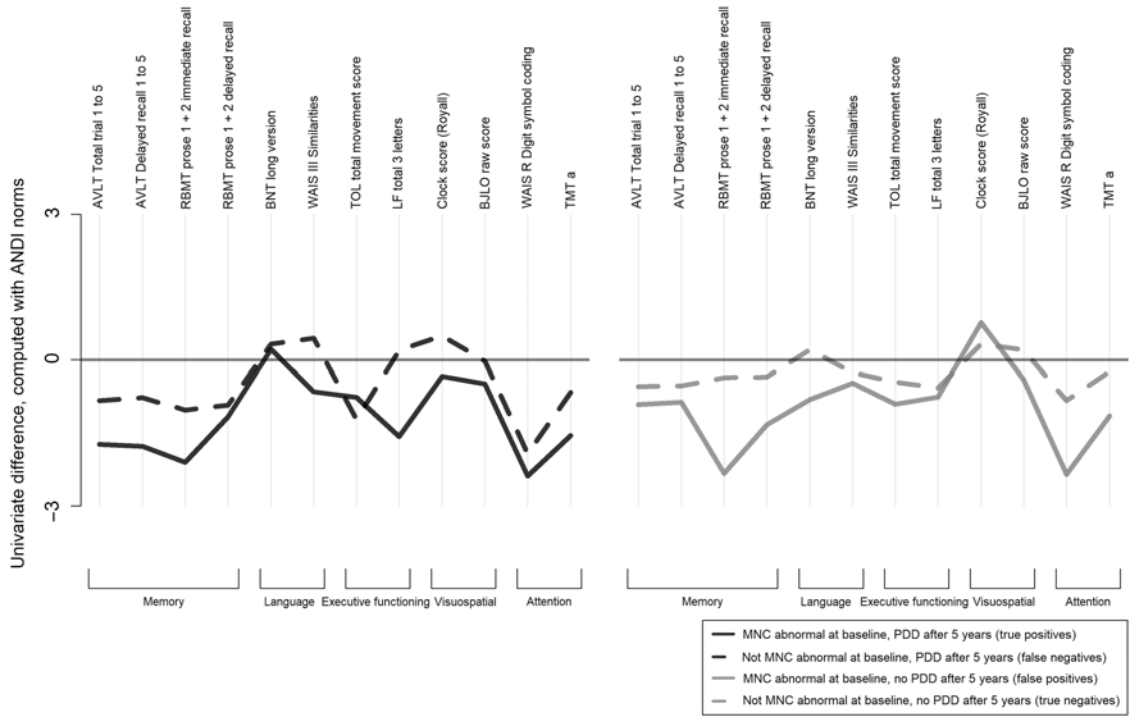

Fig. 2. Mean demographically corrected z scores for PD patients at baseline. Black lines indicate PD patients who progressed to PDD after 5 years. Gray lines indicate no PDD after 5 years. The solid line is MNC abnormal, and dashed is not MNC abnormal.

sensitivity and specificity were higher for the multivariate method than for the two univariate methods.

Cognitive Domains

We explored which cognitive domains were most often impaired in PD patients who were MNC abnormal, and whether there was a distinct profile for the patients who progressed to PDD. Figure 2 shows the mean demographically corrected z scores at baseline. Negative z scores indicate worse performance than the norm. From the figure, it can be observed that those who were MNC abnormal at baseline (solid lines) mainly showed impairment on the Rivermead Behavioural Memory Test and were slightly more impaired on the Trail Making Test (TMT) part A and the Wechsler Adult Intelligence Scale Revised edition (WAIS-R) Digit Symbol Coding task compared to those who were not MNC abnormal at baseline (dashed lines). Performance on the WAIS-R Digit Symbol Coding task seemed to be low for both those who were MNC normal and MNC abnormal at baseline. This is probably due to PD pathology affecting motor performance.

No clear difference is visible between those who progressed to PDD after 5 years (black lines), and those who did not (gray lines). From the figure, we can see that the MNC abnormality is primarily driven by the Auditory Verbal Learning Test (AVLT), Rivermead Behavioral Memory Test (RBMT), and letter fluency variables, as this is where the deepest troughs are for the MNC abnormal group (solid lines). If we look at just these tests, those who eventually do develop PDD (black solid line) are those with low scores on AVLT and letter fluency. Therefore, these tests seem to be most sensitive. The figure in Supplement 1 plots a line for every individual patient, and thus provides more detailed information on individual differences.

DISCUSSION

We investigated three methods for detecting cognitive abnormalities in PD patients that predict progression to PDD. We compared the predictive performance of the PD-MCI criteria, applied either with traditional normative data (Broeders et al., 2013) or with the ANDI normative database, to the performance of MNC using the ANDI database. We found that the number of patients diagnosed with PD-MCI at baseline differed between these methods. The original PD-MCI criteria as applied by Broeders et al. (2013) resulted in 35% of the PD patients being diagnosed with PD-MCI. Using the same criteria but with ANDI normative data, this decreased to 27%. The MNC method applied with ANDI concluded that 26% of the patients were cognitively abnormal at baseline. In the literature, the frequency with which cognitive impairments in PD patients are reported differs greatly between studies (probably due to differences in methodology and in sample characteristics, such as disease duration or severity). Studies with comparable methods to ours (1.5 SD deviating on at least 2 out of 10 tests) show that between 21% and 60.5% of PD patients are diagnosed as PD-MCI (Domellöf et al., 2015; Galtier et al., 2016; Gasca-Salas et al., 2014; Hobson & Meara, 2015; Janvin et al., 2003; Pedersen et al., 2017; Santangelo et al., 2015). The new MNC technique yields a number that lies at the low end of this range.

In terms of prediction of progression to PDD, the MNC method applied with the ANDI database performed best. Although classification accuracy was not statistically different between methods, both sensitivity and specificity were higher for the multivariate method than for the two PD-MCI criteria methods. This suggests that the improvement is mainly due to use of a multivariate statistical technique and not due to use of a large aggregated database. This was true for both the prediction of PDD after 3 and after 5 years. Between the two PD-MCI criteria methods, there was little difference in terms of accuracy. The PD-MCI criteria applied with ANDI resulted in a slightly lower sensitivity and a slightly higher specificity compared to the PD-MCI criteria as applied with Broeders et al. (2013). Just using the ANDI database instead of traditional norms therefore does not seem to improve prediction.

Previous studies that also used 1.5 SD as a cutoff score reported a sensitivity of the PD-MCI criteria for PDD ranging from .52 (Pedersen et al., 2017) to .92 (Gasca-Salas et al., 2014) and specificity ranging from .46 (Galtier et al., 2016) to .94 (Hobson & Meara, 2015). Therefore, the sensitivity and specificity estimates obtained with the MNC are at the high end of the spectrum.

For all three methods studied in this article, a decrease in sensitivity can be observed between the 3-year and 5-year follow-ups. An explanation would be that with a short period between baseline and PDD diagnosis, most patients who progressed to dementia were already rather severely impaired, leading to a high sensitivity. With more time between baseline and PDD diagnosis, some patients who progressed to dementia may have been unimpaired at baseline, leading to a lower sensitivity. Similarly, a small increase in specificity between the 3-year and 5-year follow-ups can be observed. This is explained by the time it takes to progress to dementia: patients who are impaired at baseline may still not progress to dementia in the first few years after baseline, leading to a low specificity. As more time passes however, patients who were impaired at baseline will probably progress to dementia, leading to an increase in specificity.

In our sample, it appeared that the patients who were MNC abnormal at baseline mainly experienced difficulties in memory and attention. These findings are in line with earlier studies, where PD-MCI patients showed problems with memory (Aarsland et al., 2009; Galtier et al., 2016; Janvin et al., 2003; Mamikonyan et al., 2009; Pedersen et al., 2017) and attention (Elgh et al., 2009; Mamikonyan et al., 2009; Pedersen et al., 2017). Difficulties in other domains, such as executive functions (Elgh et al., 2009; Janvin et al., 2003; Pedersen et al., 2017), language (Hobson et al., 2004), and visuospatial abilities (Janvin et al., 2003; Williams-Gray et al., 2007), are also reported in the literature but were not found in our sample. This could be due to the heterogeneity of the PD population. In our patient sample, newly diagnosed PD patients were examined. Since disease duration determines to a large extent which cognitive functions are impaired (Hobson et al., 2004; Hughes et al., 2000; Litvan et al., 2011), our patients might eventually also develop difficulties in these other cognitive domains (Muslimović et al., 2009). One difficulty with studying cognition in a movement disorder like PD is that motor problems may present confound, as many of the tests used to examine cognition require motor skills to some extent. In our sample, patients with MCI also had more severe motor problems [as evidenced by their Unified Parkinson's Disease Rating Scale (UPDRS) and Hoehn & Yahr scores]. They are indeed somewhat impaired on tests that require motor skills, but they are equally impaired on tests that do not require motor skills, like memory tests. Therefore, we argue that MCI is primarily a reflection of cognitive problems, not motor problems, in this sample.

There are several limitations to our study. First, the number of patients was not very large ($n = 123$) and loss to follow-up was quite high (21% at 3 years, and another 25% at 5 years). However, the numbers lost to follow-up are not different between those cognitively normal or abnormal at baseline (in the tables in Supplement 2 a specification of which patients were lost to follow-up is given). Because a formal test of a difference between rates requires a very large sample size if the prevalence of a disease is low (Carley et al., 2005), the power to detect differences between the sensitivity and specificity of the different methods was low, and confidence intervals were broad. We recommend that future studies either collect a much larger sample, or find a way to synthesize the literature to obtain a better estimate of these rates. Second, the age range in the current sample is quite broad (35–84 years at baseline). Consequently, the sample may have been heterogeneous with regard to cognitive functioning. Third, even though all cognitive domains prescribed by the PD-MCI criteria (Litvan et al., 2012) were studied, tests outside the traditional domains of neuropsychology have been shown to be informative in this type of clinical research. For example, it is known that decision-making in PD patients is impaired (Kobayakawa et al., 2008; Mimura et al., 2006). Therefore, future research could also include other domains of cognition when evaluating PD patients, to better represent the impairments with which they struggle.

In diagnosing PDD, we applied the MDS consensus criteria (Dubois et al., 2007; Emre et al., 2007). This procedure did not entail an in-depth cognitive assessment, which is a limitation. Moreover, we used the MMSE which, since the start of the study in 2003, has been criticized and is no longer considered the optimal choice for this purpose (Hoops et al., 2009; Skorvanek et al., 2018). For future research, where cognition needs to be assessed in a time-constrained setting, we would advise to use other instruments such as the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005), or the Disease-Cognitive Rating Scale (PD-CRS) (Pagonabarraga et al., 2008, p. 14).

In this study, no correction was made for premorbid IQ, even though IQ influences the scores that patients achieve on neuropsychological tests (Testa et al., 2009). Corrections for premorbid cognitive functioning were not performed here, because they were not included in the original determination of PD-MCI by Broeders et al. (2013), and were not available

from the ANDI database. We recommend that future work will take premorbid IQ in account, for example, by correcting scores using a reading test like the reading subtest of the Wide Range of Achievement Test in Autism (WRAT) (Wilkinson & Robertson, 2006) or the Test of Premorbid Functioning (TOPF) (Pearson Assessment, 2009). Furthermore, in this study, we have focused on just PD and progression to PDD. For future studies, we recommend that the influence of more disorders, such as Parkinson's plus disorders and psychiatric disorders, is taken into account.

Subjective complaints were used in PD-MCI criteria and MNC. Therefore, subjective complaints played a large role in determining the diagnoses in this study, while they were established using only two questions. Possibly, higher specificity and sensitivity would have been obtained, had we established subjective complaints more formally, for example, with a longer, validated questionnaire, ideally including reports by relatives, caregivers, and clinicians. Instead, for patients without subjective complaints, deviation on at least four neuropsychological tests was used as a criterion for PD-MCI, and a stricter criterion was used for MNC.

It is not easy to apply MNCs in daily clinical practice. We therefore developed the user-friendly ANDI website. Currently, ANDI is only applicable to the Dutch-speaking population. However, we provide it as an open source infrastructure. It is possible to copy ANDI and recreate the system in other countries, which would only require a local (aggregated) normative sample. Doing so would make MNCs easier applicable in the clinical setting.

In sum, using the large aggregated data set of the ANDI database circumvents limitations found in standard norm-referenced testing. The MNC method works at least as well as the more conventional approach of the PD-MCI criteria and trended toward a better classification accuracy for predicting future dementia.

ACKNOWLEDGEMENTS

We have no conflicts of interest to disclose. This study was funded by The Netherlands Organization for Scientific Research (NWO) (project number 480-12-015).

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S1355617719000298

REFERENCES

Aarsland, D., Andersen, K., Larsen, J.P., Lolk, A., Nielsen, H., & Kragh–Sørensen, P. (2001). Risk of dementia in Parkinson's disease: A community-based, prospective study. *Neurology*, *56*(6), 730–736. doi: 10.1212/WNL.56.6.730.

Aarsland, D., Brønnick, K., Larsen, J.P., Tysnes, O.B., Alves, G., & Norwegian ParkWest Study Group. (2009). Cognitive impairment in incident, untreated Parkinson disease: The Norwegian ParkWest Study. *Neurology*, *72*(13), 1121–1126.

Agelink van Rentergem, J.A., de Vent, N.R., Schmand, B.A., Murre, J.M.J., & Huizenga, H.M. (2017a). Multivariate normative comparisons for neuropsychological assessment by a multilevel factor structure or multiple imputation approach. *Psychological Assessment*. Advance online publication.

Agelink van Rentergem, J.A., Murre, J.M.J., & Huizenga, H.M. (2017b). Multivariate normative comparisons using an aggregated database. *PLoS ONE*, *12*, 1–18.

Agresti, A. & Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119–126.

Benton, A.L. & Hamsher, K. (1983). *Multilingual Aphasia Examination*. Iowa City: AJA Associates.

Benton, A.L., Hamsher, K., Varney, N., & Spreen, O. (1983). *Contributions to Neuropsychological Assessment – A Clinical Manual*. New York: Oxford University Press.

Binder, L.M., Iverson, G.L., & Brooks, B.L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*(1), 31–46. doi: 10.1093/arclin/acn001.

Broeders, M., De Bie, R.M.A., Velseboer, D.C., Speelman, J.D., Muslimovic, D., & Schmand, B. (2013). Evolution of mild cognitive impairment in Parkinson disease. *Neurology*, *81*(4), 346–352.

Carley, S., Dosman, S., Jones, S.R., & Harrison, M. (2005). Simple nomograms to calculate sample size in diagnostic studies. *Emergency Medicine Journal*, *22*(3), 180–181.

Caviness, J.N., Driver-Dunckley, E., Connor, D.J., Sabbagh, M.N., Hentz, J.G., Noble, B., Evidente, V.G.H., Shill, H.A., & Adler, C.H. (2007). Defining mild cognitive impairment in Parkinson's disease. *Movement Disorders*, *22*(9), 1272–1277. doi: 10.1002/mds.21453.

Collin, C., Wade, D.T., Davies, S., & Horne, V. (1988). The Barthel ADL Index: A reliability study. *International Disability Studies*, *10*(2), 61–63. doi: 10.3109/09638288809164103

Crawford, J.R. & Garthwaite, P.H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40*, 1196–1208. doi: 10.1016/S00283932(01)00224-X.

Culbertson, W.C. & Zillmer, E.A. (1998). The Tower of London DX: A standardized approach to assessing executive functioning in children. *Archives of Clinical Neuropsychology*, *13*(3), 285–301. doi: 10.1016/S0887-6177(97)00033-4.

de Vent, N.R., Agelink van Rentergem, J.A., Schmand, B.A., & Murre, J.M.J., ANDI Consortium, & Huizenga, H.M. (2016) Advanced Neuropsychological Diagnostics Infrastructure (ANDI): A normative database created from control datasets. *Frontiers in Psychology*, *7*(1601), 1–10. doi: 10.3389/fpsyg.

Domellöf, M.E., Ekman, U., Forsgren, L. & Elgh, E. (2015). Cognitive function in the early phase of Parkinson's disease, a five-year follow-up. *Acta Neurologica Scandinavica*, *132*(2), 79–88.

Dubois, B., Burn, D., Goetz, C., Aarsland, D., Brown, R.G., Broe, G.A., Dickson, D., Duyckaerts, C., Cummings, J., Gauthier, S., & Korczyn, A. (2007). Diagnostic procedures for Parkinson's disease dementia: Recommendations from the movement disorder society task force. *Movement Disorders*, *22*(16), 2314–2324.

Elgh, E., Domellöf, M., Linder, J., Edström, M., Stenlund, H., & Forsgren, L. (2009). Cognitive function in early Parkinson's disease: A population-based study. *European Journal of Neurology*, *16*(12), 1278–1284.

Emre, M., Aarsland, D., Brown, R., Burn, D.J., Duyckaerts, C., Mizuno, Y., Broe, G.A., Cummings, J., Dickson, D.W., Gauthier, S., & Goldman, J. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders*, *22*(12), 1689–1707. doi: 10.1002/mds.21507.

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*(3), 189-198. doi: 10.1016/0022-3956(75)90026-6.

Galtier, I., Nieto, A., Lorenzo, J.N., & Barroso, J. (2016). Mild cognitive impairment in Parkinson's disease: Diagnosis and progression to dementia. *Journal of Clinical and Experimental Neuropsychology*, *38*(1), 40–50.

Gasca-Salas, C., Estanga, A., Clavero, P., Aguilar-Palacio, I., González-Redondo, R., Obeso, J.A., & Rodríguez-Oroz, M.C. (2014). Longitudinal assessment of the pattern of cognitive decline in non-demented patients with advanced Parkinson's disease. *Journal of Parkinson's Disease*, *4*(4), 677–686. doi: 10.3233/JPD-140398.

Hobson, P. & Meara, J. (2004). The risk and incidence of dementia in a cohort of older subjects with Parkinson's disease int the UK. *Movement Disorders*, *19*(9), 1043–1049. doi: 10.1002/mds.20216.

Hobson, P. & Meara, J. (2015). Mild cognitive impairment in Parkinson's disease and its progression onto dementia: A 16-year outcome evaluation of the Denbighshire cohort. *International Journal of Geriatric Psychiatry*, *30*(10), 1048–1055. doi: 10.1002/gps.4261.

Hoogland, J., Boel, J.A., de Bie, R.M., Geskus, R.B., Schmand, B.A., Dalrymple-Alford, J.C., Marras, C., Adler, C.H., Goldman, J.G., Tröster, A.I., & Geurtsen, G.J. (2017). Mild cognitive impairment as a risk factor for Parkinson's disease dementia. *Movement Disorders*, *32*(7), 1056–1065. doi: 10.1002/mds.27002.

Hoops, S., Nazem, S., Siderowf, A.D., Duda, J.E., Xie, S.X., Stern, M.B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, *73*(21), 1738–1745.

Hughes, T.A., Ross, H.F., Musa, S., Bhattacherjee, S., Nathan, R.N., Mindham, R.H.S., & Spokes, E.G. (2000). A 10-year study of the incidence of and factors predicting dementia in Parkinson's disease. *Neurology*, *54*(8), 1596–1603.

Huizenga, H.M., Smeding, H., Grasman, R.P., & Schmand, B. (2007). Multivariate normative comparisons. *Neuropsychologia*, *45*, 2534–2542. doi: 10.1016/j.neuropsychologia.2007.03.011.

Janvin, C., Aarsland, D., Larsen, J.P., & Hugdahl, K. (2003). Neuropsychological profile of patients with Parkinson's disease without dementia. *Dementia and Geriatric Cognitive Disorders*, *15*(3), 126–131. doi: 10.1159/000068483.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia, PA: Lea & Febiger.

Kobayakawa, M., Koyama, S., Mimura, M., & Kawamura, M. (2008). Decision making in Parkinson's disease: Analysis of behavioral and physiological patterns in the Iowa gambling task. *Movement Disorders*, *23*(4), 547–552.

Lezak, M.D., Howieson, D.B., Bigler, E.D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). New York: Oxford University Press.

Litvan, I., Aarsland, D., Adler, C.H., Goldman, J.G., Kulisevsky, J., Mollenhauer, B., Rodriguez-Oroz, M.C., Tröster, A.I., & Weintraub, D. (2011). MDS task force on mild cognitive impairment in Parkinson's disease: Critical review of PDMCI. *Movement Disorders*, *26*(10), 1814–1824. doi: 10.1002/mds.23823.

Litvan, I., Goldman, J.G., Tröster, A.I., Schmand, B.A., Weintraub, D., Petersen, R.C., Mollenhauer, B., Adler, C.H., Marder, K., Williams-Gray, C.H., & Emre, M. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Movement Disorders*, *27*(3), 349–356. doi: 10.1002/mds.24893.

Mamikonyan, E., Moberg, P.J., Siderowf, A., Duda, J.E., Ten Have, T., Hurtig, H.I., Stern, M.B., & Weintraub, D. (2009). Mild cognitive impairment is common in Parkinson's disease patients with normal Mini-Mental State Examination (MMSE) scores. *Parkinsonism & Related Disorders*, *15*(3), 226–231.

Mimura, M., Oeda, R., & Kawamura, M. (2006). Impaired decision-making in Parkinson's disease. *Parkinsonism & Related Disorders*, *12*(3), 169–175.

Muslimović, D., Post, B., Speelman, J.D., De Haan, R.J., & Schmand, B. (2009). Cognitive decline in Parkinson's disease: A prospective longitudinal study. *Journal of the International Neuropsychological Society*, *15*(3), 426–437.

Muslimović, D., Post, B., Speelman, J.D., & Schmand, B. (2005). Cognitive profile of patients with newly diagnosed Parkinson disease. *Neurology*, *65*(8), 1239–1245. doi: 10.1212/01.wnl.0000180516.69442.95.

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L. & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*(4), 313–324.

Pagonabarraga, J., Kulisevsky, J., Llebaria, G., García-Sánchez, C., Pascual-Sedano, B., & Gironell, A. (2008). Parkinson's disease-cognitive rating scale: A new cognitive scale specific for Parkinson's disease. *Movement Disorders*, *23*(7), 998–1005.

Pearson Assessment (2009). *Advanced Clinical Solutions for the WAIS-IV/WMS-IV*. San Antonio, TX: Author.

Pedersen, K.F., Larsen, J.P., Tysnes, O.B., & Alves, G. (2017). Natural course of mild cognitive impairment in Parkinson disease: A 5-year population-based study. *Neurology*, *88*(8), 767–774.

Reitan, R.M. (1992). *Trail Making Test: Manual for Administration and Scoring*. Tucson, AZ: Reitan Neuropsychological Laboratory.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris, France: Presses Universitaires de France.

Royall, D.R., Cordes, D.A., & Polk, M. (1998). CLOX: An executive clock drawing task. *Journal of Neurology, Neurosurgery, and Psychiatry*, *64*, 588–594.

Santangelo, G., Vitale, C., Picillo, M., Moccia, M., Cuoco, S., Longo, K., Pezzella, D., di Grazia, A., Erro, R., Pellecchia, M.T., & Amboni, M. (2015). Mild cognitive impairment in newly diagnosed Parkinson's disease: A longitudinal prospective study. *Parkinsonism & Related Disorders*, *21*(10), 1219–1226.

Schwab, J.F. & England, A.C. (1969). Projection technique for evaluating surgery in Parkinson's disease. In: F.H. Billingham & M.C. Donaldson (Eds.), *Third Symposium on Parkinson's Disease* (pp. 152–157). Edinburgh: Churchill Livingstone.

Skorvanek, M., Goldman, J.G., Jahanshahi, M., Marras, C., Rektorova, I., Schmand, B., van Duijn, E., Goetz, C.G., Weintraub, D., Stebbins, G.T., & Martinez-Martin, P. (2018). Global scales for cognitive screening in Parkinson's disease: Critique and recommendations. *Movement Disorders*, *33*(2), 208–218.

Su, T., Schouten, J., Geurtsen, G.J., Wit, F.W., Stolte, I.G., Prins, M., Portegies, P., Caan, M.W., Reiss, P., Majoie, C.B., & Schmand, B.A. (2015). Multivariate normative comparison, a novel method for more reliably detecting cognitive impairment in HIV infection. *AIDS*, *29*(5), 547–557.

Testa, S.M., Winicki, J.M., Pearlson, G.D., Gordon, B. & Schretlen, D.J. (2009). Accounting for estimated IQ in neuropsychological test performance with regression-based techniques. *Journal of the International Neuropsychological Society*, *15*(6), 1012–1022.

Van der Putten, J.J., Hobart, J.C., Freeman, J.A., & Thompson, A.J. (1999). Measuring change in disability after inpatient rehabilitation: Comparison of the responsiveness of the Barthel index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery & Psychiatry*, *66*(4), 480–484.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised (WAIS-R)*. New York: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale (WAIS-III)* (3rd ed.). New York: Psychological Corporation.

Wilkinson, G.S. & Robertson, G.J. (2006). *Wide Range Achievement Test* (4th ed.). Lutz, FL: Psychological Assessment Resources.

Williams-Gray, C.H., Foltynie, T., Brayne, C.E.G., Robbins, T.W. & Barker, R.A. (2007). Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. *Brain*, *130*(7), 1787–1798. doi: 10.1093/brain/awm111.

Wilson, B., Cockburn, J., & Baddeley, A. (1983). *Rivermead Behavioural Memory Test*. Reading, UK: Thames Valley Test Company.

Zachary, R.A. & Gorsuch, R.L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, *41*(1), 86–94.

Zigmond, A.S., & Snaith, R.P. (1983). The hospital anxiety and depression scale. *Acta psychiatrica scandinavica*, *67*(6), 361–370.