Course of cognitive decline in Parkinson's disease: A meta-analysis

DINO MUSLIMOVIĆ,[1] BEN SCHMAND,[1,2] JOHANNES D. SPEELMAN,[1]
AND ROB J. DE HAAN[3]

[1]Department of Neurology, Academic Medical Centre, University of Amsterdam, The Netherlands
[2]Department of Psychonomics, University of Amsterdam, The Netherlands
[3]Department of Clinical Epidemiology and Biostatistics, Academic Medical Centre, University of Amsterdam, The Netherlands

Abstract

A meta-analysis was conducted on 25 longitudinal studies involving 901 initially non-demented Parkinson's disease (PD) patients to examine the magnitude of decline across multiple cognitive domains associated with disease progression. Pooled effect sizes reflecting the standardized difference between baseline and follow-up neuropsychological performance were calculated for 8 cognitive domains using a random-effects model. Relatively small effect sizes were found across all cognitive domains ($d = .00 - .40$). During a mean follow-up interval of 29 months, significant declines were detected in global cognitive ability ($d = .40$), visuoconstructive skills ($d = .32$), and memory ($d = .29$). Age showed a significant relation with decline in global cognitive ability and memory. Lower educational level was associated with greater decline in all cognitive domains. Studies with longer follow-up intervals yielded larger effect sizes for global cognitive ability. In non-demented PD patients, changes in cognitive functions over time appear to be modest. Educational level, age, and length of the follow-up interval are likely to affect the magnitude of decline in several domains. Methodological flaws, such as selection bias and uncontrolled practice effects, may have caused underestimation of the true extent of decline (*JINS*, 2007, *13*, 920–932.)

Keywords: Parkinson's disease, Cognitive decline, Neuropsychological tests, Longitudinal studies, Prognosis, Meta-analysis

INTRODUCTION

There is now ample evidence that patients with Parkinson's disease (PD), even in the absence of clinically apparent dementia, can exhibit selective impairments in several cognitive domains (Dubois & Pillon, 1997). Some of these impairments, such as deficits in executive function and memory are observed already in the early stages of the disease (Muslimovic et al., 2005), whereas others, such as deficits in visuospatial abilities seem to occur later in the disease process (Della Sala et al., 1986; Levin et al., 1991). Current knowledge about cognitive changes in PD is largely obtained from cross-sectional studies of patients at various stages of the disease. Whereas cognitive decline is considered to be an integral part of the natural course of PD, there is little information regarding the exact pattern and degree of changes in specific cognitive domains over time. Most longitudinal studies examining non-demented PD patients have focused on the incidence and predictive factors of dementia rather than on the nature of cognitive deterioration itself.

Studies addressing the progression of cognitive deficits in PD have produced varying results with respect to the affected functional domains and the severity of the reported changes. For example, Bayles et al. (1996) reported significant deterioration in global cognitive ability among a subset of PD patients two years after the initial assessment, whereas another study described fairly stable performance even after four years (Growdon et al., 1990). Similarly, executive functions have been found by some to deteriorate (Caparros-Lefebvre et al., 1995) and by others to remain stable (Ramirez-Ruiz et al., 2005). These discrepant findings may be explained in part by differences in the methods of patient selection, the cognitive tests employed, and by the generally small sample sizes, which might have reduced

Correspondence and reprint requests to: D. Muslimović, AMC Neurology, H2-222, PO Box 22660, 1100 DD Amsterdam, The Netherlands. E-mail: d.muslimovic@amc.uva.nl

the statistical power to detect subtle changes. Moreover, some studies do not explicitly distinguish between demented and non-demented patients at the inclusion in the study, using instead an unselected population. In light of these inconsistencies and limitations, we felt that a meta-analysis of the published data was needed to clarify the pattern and extent of cognitive changes that accompany the progression of PD.

The primary aim of this study was to determine the magnitude and pattern of cognitive decline in initially non-demented PD patients by quantitatively combining data from previous studies using meta-analytic techniques. Studies included in this review adopted a longitudinal design, and the outcome measure in the current analysis represents the difference between patient's cognitive performance at initial evaluation when no clinical indications of dementia were present and their own scores at follow-up assessment. The neuropsychological tests were grouped into functional domains to aid in the interpretation of data. In addition, we evaluated the influence of demographic variables, such as age and education and clinical characteristics such as disease duration on the magnitude of cognitive changes. Furthermore, we assessed the effect of the length of follow-up.

METHODS

The data included in this study were obtained in compliance with the guidelines of the Helsinki Declaration.

Literature Search

A comprehensive literature search of Medline (1966–2006), PsychInfo (1972–2006) and Biological Abstracts (1993–2006) was conducted to identify articles for inclusion in the review. The keywords used in the search strategy were *Parkinson disease*, or *Parkinson's disease* in combination with *cognition*, or *cognitive impairment*, or *memory*, or *executive function*, or *neuropsychological tests*, and *longitudinal studies*, or *prognosis*, or *progression*. The search was completed in January 2006 and was limited to English-language articles. Articles were inspected for relevant references to locate additional studies for inclusion in the review.

Inclusion Criteria

To be included in the analysis, studies had to meet the following criteria:

- The diagnosis of idiopathic PD had to be made according to validated clinical criteria (e.g. United Kingdom Parkinson's Disease Society Brain Bank criteria [UKPDSBBC]; Gib & Lees, 1988). Studies, which did not explicitly report diagnostic criteria were still considered for inclusion if the diagnosis was made or confirmed by a board-certified neurologist.

- Patients had to be examined prospectively on at least two occasions with the same neuropsychological tests (i.e., a longitudinal design).

- Patients had to be free of clinical dementia at initial evaluation. Patients were considered to be demented if their diagnoses were made according to standardized clinical criteria (e.g., Diagnostic and Statistical Manual of Mental Disorders [DSM]; American Psychiatric Association, 1994), or if their performance fell below a traditionally employed cutoff on a screening measure for dementia. If the study sample comprised both demented and non-demented patients at the initial evaluation, the results of non-demented patients had to be analyzed separately in the original article in order to be included.

- At least one standardized neuropsychological test had to be employed as a dependent variable.

- Test scores had to be presented for the PD group both at baseline and at follow-up (mean and standard deviations), or other statistics had to be reported that could be converted to effect sizes (e.g. exact p-values or t values). In case this information was not originally reported, we contacted the study authors to obtain the relevant statistics.

In cases where different papers reported data concerning the same group of patients, the study with the largest sample or with the longest follow-up period (provided that the difference in the number of participants was not substantial) was included in the analysis.

Exclusion Criteria

Reports published only in abstract form were excluded. Clinical trials were excluded for two reasons. First, patients entered into trials may not be representative of the population with the disorder (Laupacis et al., 1994). Second, we wanted to examine the *natural* course of cognitive decline in PD. However, studies that included a control sample of PD patients who were assessed serially but who did not undergo an experimental treatment (neither active nor placebo) during the test-retest interval, were considered eligible. If the study sample comprised PD patients with evidence of delirium either at baseline or at follow-up, the results obtained from these patients were not included in the meta-analysis.

Outcome Measures

A number of cognitive domains were assessed across studies and many different tests were used to measure these domains. Classification of tests into functional domains was based on the detailed descriptions of task characteristics and the corresponding area of cognitive functioning described in the two standard textbooks of neuropsychological assessment, which are widely accepted both in clinical practice and in cognitive research (Lezak et al., 2004; Strauss et al., 2006). Tests from the individual studies were categorized into the following eight functional domains: global

cognitive ability, verbal ability, memory, verbal fluency, mental flexibility and reasoning, attention and processing speed, visuoperceptual functions, and visuoconstructive skills. Appendix A lists the tests that were included in each cognitive domain.

Calculation of Effect Sizes and Statistical Analysis

From the data reported in each study, the effect-size estimate Hedge's g (Hedges & Olkin, 1985) was calculated, indicating the mean difference between baseline and follow-up divided by the pooled standard deviation. When means and standard deviations were not available, effect sizes were calculated from other reported statistics using the methods described by Rosenthal (1991). Cognitive test results obtained at different measurement points are expected to be dependent, and effect sizes need to be corrected for this correlation. However, none of the studies in the current meta-analysis reported the correlation between baseline and follow-up test scores. Therefore, these correlation coefficients were estimated from a few studies that provided raw scores, *t*-tests, or corresponding *p*-values (Azuma et al., 2003; Bayles et al., 1996; Caparros-Lefebvre et al., 1995; Katsarou et al., 1998; Ramirez-Ruiz et al., 2005). These studies employed follow-up intervals similar to the average follow-up period of the review (see Table 1). Separate correlation coefficients were calculated for each cognitive domain. Subsequently, these correlation coefficients were used to adjust the variance (V) for repeated measurements using the formula:

$$V = \frac{(s_b^2 + s_f^2 - 2rs_b s_f)}{n(s_b^2 + s_f^2)/2}$$

where s_b and s_f are the standard deviations at baseline and follow-up, respectively, r is the correlation coefficient between baseline and follow-up test scores, and n is the sample size.

If no exact values were reported for non-significant results, we adopted a conservative approach, and set the effect size for that measure at zero (Rosenthal, 1991). This method was applied for four outcome measures derived from two studies (Growdon et al., 1990; Hovestadt, 1990). If the study reported a range of the significance level rather than exact values, the *p*-value used to calculate the effect size was set only marginally lower (i.e., for $p < .05$ it was set at .049, for $p < .01$ at .0099, and for $p < .001$ at .00099). This was the case in only one study (Caparros-Lefebvre et al., 1995).

From the effect sizes obtained in individual studies (i.e., g-values), a combined *d*-value was calculated, expressing the magnitude of cognitive decline across studies. This *d*-value was weighted for the sample sizes of the individual studies in order to correct for upwardly biased estimation of the effect in small samples (Hedges & Olkin, 1985). Separate pooled effect sizes were calculated for each of the eight cognitive domains. In addition, 95% confidence intervals (CI) were computed around each mean weighted effect size to provide an estimate of the variability of *d* and to test whether the *d*-value was significantly different from zero.

Because the studies included in the present analysis are diverse with regard to both clinical variables and methods of assessment of cognitive functions, heterogeneity in their results is to be expected. Therefore, a random-effects model was considered appropriate to generate the pooled effect sizes (Raudenbush, 1994).

For both individual and pooled effect sizes, a positive direction indicated worsening of cognitive performance over time. In accordance with Cohen's guidelines (1988) for interpreting effect sizes, values between .20 and .50 were considered as small, values between .50 and .80 as medium, and values above .80 as large.

In studies that used more than one measure to assess a particular cognitive domain, an averaged effect size was calculated for the final analysis. Thus, each study was allowed to contribute only one effect size for each functional domain. This strategy was used to avoid that one study dominated the results.

In studies in which the sample was divided in subgroups on the basis of certain variables (e.g., presence *vs.* absence of depression, Hoehn & Yahr score), effect sizes were calculated from the data of the whole group. If such findings were not reported, the effect sizes were calculated for each subgroup separately and a weighted average of these effect sizes was used in the final analysis.

To determine whether there was variation in effect sizes across studies, the chi-square statistic Q was calculated. The Q-statistic quantifies the degree to which the studies contributing to each respective mean effect size can be regarded as homogeneous. A significant Q-value indicates heterogeneity among studies contributing to the particular mean, in which case a further search for potential moderator variables is needed. It is recommended, however, to perform additional analyses even when the Q-statistic is not significant, to determine whether the magnitude of the effect sizes covaries with some attributes of the study (Rosenthal, 1991).

Because the Q-statistic is believed to have limited value in detecting true heterogeneity among studies because of its rather low power, particularly when meta-analyses encompass a relatively small number of studies, an alternative index I^2 indicating the degree of inconsistency in the study results was calculated using the equation $I^2 = (Q - df)/Q \times 100\%$, where Q is Cochran's heterogeneity statistic and df the degrees of freedom (the number of studies -1). The I^2 index reflects the percentage of total variation across studies that are caused by heterogeneity rather than chance (Higgins et al., 2003). A value of 0% indicates no observed heterogeneity and larger values show increasing heterogeneity. In the moderator analysis, the Qw statistic indicates the degree of heterogeneity of studies within categories. The Q_B statistic refers to a difference between categories of the moderator variable, and if

statistically significant, suggests the influence of this moderating variable. All analyses were performed using the statistical software MetaWin version 2.0 (Rosenberg et al., 2000).

Publication Bias

Extracting data from published studies may bias results in favor of a significant mean effect size, because studies with non-significant results are less likely to be published (Rosenthal, 1991). To examine the possibility of publication bias, the fail-safe N was calculated for each mean weighted effect size to estimate the number of negative studies which would be necessary to render the results non-significant (Rosenthal, 1991). In addition, Rosenthal (1991) recommends the value of $5k + 10$ (where k is the number of studies used to calculate the effect size) as a reasonable, conservative estimate of existing unpublished or unretrieved studies against which to test a fail-safe calculation. If the fail-safe N is large relative to this estimated number of negative studies one could be fairly confident that the observed result is a reliable estimate of the true effect.

Moderator Variables

We assessed the potential influence of demographic and clinical variables on effect sizes using categorical models. The demographic variables included age of the patients and years of education. Among clinical variables, only duration of disease was included in the analysis. The potential influence of other demographic and clinical characteristics (e.g. sex, severity of disease, medication) on effect sizes could not be evaluated due to a rather small number of studies reporting exact information for these parameters.

With regard to study characteristics, we were particularly interested in the effect of length of follow-up period. Therefore, we conducted additional analyses to examine the extent to which changes in cognitive functioning varied as a function of follow-up interval. For all variables included in moderator analyses, the division of groups was based on a median split.

RESULTS

Studies Retrieval

The literature search identified 85 longitudinal studies on PD, excluding those that examined the effects of treatment. Of these, three reports were published in abstract form; 12 studies reported data from (almost) the same samples; 11 studies included a number of patients who either met the clinical criteria for dementia or were suspected of having dementia at the initial assessment; 24 studies did not employ neuropsychological tests as the outcome measure; one study employed an uncommon test that was not comparable with other measures; and nine studies did not report relevant

statistics to calculate effect sizes. Twenty-five of the studies that had been identified met criteria and were included in this meta-analysis. Demographic and clinical characteristics of each of these studies are shown in Table 1.

Participant and Study Characteristics

In total, neuropsychological test results from 901 initially non-demented patients with PD who underwent both baseline and follow-up assessment were recorded in the meta-analysis. As can be seen in Table 2, patients were on average 64.7 years of age at initial evaluation and were affected by the disease for approximately 8 years. The mean Hoehn and Yahr score (Hoehn & Yahr, 1967) indicated a moderate degree of motor disorder. From the studies that provided sufficient information, it appeared that the majority of patients were receiving anti-Parkinson's disease medication at baseline. Follow-up assessments ranged from 2.4 months to 8 years, with a mean test-retest interval of 29 months.

Thirteen studies employed a dementia screening measure, whereas in 12 studies a more comprehensive clinical examination was used to assess the presence of dementia at baseline. In three studies the presence of depression was an exclusion criterion both at baseline and at follow-up. The majority of studies (76%) included patients who remained non-demented at the follow-up evaluation. One study comprised non-medicated *de novo* patients at baseline, whereas the remaining studies included patients on anti-Parkinson's disease medication who were in later stages of disease. With regard to follow-up assessment, 88% of the studies utilized one follow-up measurement point, 8% of the studies used two measurement points, and in one study patients were examined ten times after the baseline evaluation.

Effect Sizes

A total of 154 effect sizes were extracted covering eight cognitive domains. Mean weighted effect sizes (*d*-values) for each domain are shown in Table 3. In five domains (verbal ability, verbal fluency, mental flexibility/reasoning, attention and processing speed, and visuoperceptual functions) effect sizes were negligible ($d \leq .10$). For global cognitive ability, visuoconstructive skills, and memory the effect sizes were small in magnitude, but statistically significant ($p < .05$). The greatest degree of decline was observed in global cognitive ability ($d = .40$).

The Q statistic indicated no significant heterogeneity for any of the effect sizes. This finding is further supported with index I^2, which revealed no heterogeneity whatsoever for seven functional domains and only small variability across studies that contributed data for global cognitive ability.

Publication Bias

For cognitive domains that showed significant decline, the fail-safe N and an estimate of unpublished, non-significant

Table 1. Characteristics of studies included in the meta-analysis

Study (year)	Sampling Method	N Included/ Total†	Age (years)	Duration PD (years)	Age at onset (years)	Education (years)	Follow-up (months)
Aarsland et al. (2004)	Community	36/80	67.6	—	—	—	96
Azuma et al. (2003)	MD clinic PD Association Neurology practices	69/80	68.9	5.7	63.2	14.8	25.8
Bayles et al. (1996)	MD clinic PD Association Neurology practices	70/103	—	—	—	—	24
Caparros-Lefebvre et al. (1995)	—	53/60	61.7	8.6	53.0	9.6	37
Dujardin et al. (2004)	—	44/50	63.0	1	62.0	—	36
Ebmeier et al. (1990)	Community	101/101‡	—	—	—	—	43
Firbank et al. (2005)	Local services	30/38	75.5	3.8	71.7	—	12.7
Growdon et al. (1990)	—	32/32	—	—	—	—	48
Hayashi et al. (1996)	—	29/29	58.5	5.7	52.8	—	25
Hovestadt et al. (1990)	Outpatient clinic	18/18	54.9	3.8	51.1	—	12
Hu et al. (2001)	—	8/8	64.0	9.5	54.5	—	23.6
Katsarou et al. (1998)	—	45/45	59.9	5.3	54.5	7.2	36
Morrison et al. (2004)	—	11/11	62.7	10.0	52.8	13.8	2.4
Palazzini et al. (1995)	Specialty clinic	50/91#	54.8	5.7	49.1	7.6	86.7
Perrine et al. (1998)	Specialty clinic	10/10	—	—	—	—	4.5
Ramirez-Ruiz et al. (2005)	Outpatient clinic	11/13	72.4	10.3	62.1	7.7	25
Serrano-Dueñas et al. (2005)	Specialty clinic	17/21	67.5	6.5	61.5	—	60
Schmand et al. (2000)	Outpatient clinic	16/18	61.1	16.3	44.8	10.9	6
Smeding et al. (2006)	Outpatient clinic	36/39	63.0	10.4	52.6	12.4	6
Soliveri et al. (2000)	—	12/18	57.7	12.1	45.6	8.6	21
Starkstein et al. (1990)	Outpatient clinic	49/70	66.9	8.9	54.8	13.0	44.5
Starkstein et al. (1992)	Outpatient clinic	74/105§	67.3	9.0	58.3	13.5	12
Stebbins et al. (2000)	—	10/10	62.4	—	—	12.9	12
Stern et al. (1998)	Community	40/40	78.5	—	—	9.2	12
Tachibana et al. (1995)	—	30/30	66.7	5.7	61.0	—	12

†Number of patients reported in the study/the total number of patients prior to exclusion due to insufficient data or loss to follow-up;
‡PD sample included 243 patients at baseline and 127 patients at follow-up, but cognitive performance was assessed in only 101 patients;
#The actual number of patients with follow-up was 61, but data from 11 patients who fulfilled the criteria for dementia at follow-up were excluded from the meta-analysis since there was evidence of delirium in 72% of these patients;
§The actual number of patients with follow-up was 92, but data from a subgroup of patients with major depression (N = 18) are excluded from the meta-analysis as the mean MMSE score at baseline assessment indicated presence of global cognitive decline.
MD Movement Disorder.

Table 2. Summary of demographic and clinical characteristics of studies included in the meta-analysis

Variables	M	SD	Range	N
Sample size	37.9	25.5	8–101	25
Age	64.7	6.1	54.8–78.5	21
Percentage of men	63.3	14.7	36.7–90.9	15
Education (years)	10.9	2.7	7.2–14.8	13
Duration of disease (years)	7.7	3.6	1–16.3	18
Age at onset of disease	57.0	7.0	44.8–71.8	18
Hoehn & Yahr scale	2.6	0.5	1.5–3.3	13
Percentage of patients on medication	90.9	24.7	0–100	18
Follow-up interval (months)	28.6	24.0	2.4–96	25

N = Number of studies in which the variable was reported.

studies were calculated (Table 3). The fail-safe N of 473 for global cognitive ability exceeded the estimate of 105 existing unpublished studies reporting non-significant results (tolerance level), which suggests that the observed effect cannot be explained by publication bias. Similarly, the fail-safe N of 78 studies for memory domain exceeded the tolerance level of 70, whereas only 30 studies (below the tolerance level of 45) would be required to reduce the findings regarding visuoconstructive skills to a non-significant level.

Influence of Moderator Variables

Older age at baseline was related to a greater degree of deterioration on measures of global cognitive ability and

Table 3. Pooled weighted effect sizes, confidence intervals and heterogeneity statistics for each domain of cognitive functioning

Domain	k	N patients	Nd	d	95% CI	Q	p (Q)	I^2	Fail-safe N	Tolerance level‡
Global cognitive ability	19	702	21	0.40*	0.20–0.61	19.65	.35	8	473	105
Verbal ability	8	229	12	0.07	−0.14–0.28	6.28	.51	0†	—	—
Memory	12	341	31	0.29*	0.06–0.58	9.22	.60	0†	78	70
Verbal fluency	12	320	22	0.05	−0.15–0.25	10.47	.49	0†	—	—
Mental flexibility/Reasoning	12	325	17	0.10	−0.17–0.36	9.44	.58	0†	—	—
Attention and processing speed	9	193	31	−0.01	−0.39–0.38	4.33	.83	0†	—	—
Visuoperceptual functions	8	174	10	0.05	−0.20–0.30	6.47	.49	0†	—	—
Visuoconstructive skills	7	217	10	0.32*	0.12–0.53	5.82	.44	0†	30	45

k = number of studies; Nd = number of effect sizes; d = mean weighted effect size; CI = confidence interval; Q = within domain heterogeneity; p (Q) = p value for heterogeneity; I^2 = percentage of heterogeneity due to study differences [$(Q − df)/Q \times 100\%$].
‡Tolerance level (i.e., estimated number of existing unpublished studies, 5k + 10; Rosenthal, 1991).
†Negative I^2 values are set equal to zero as recommended by Higgins et al. (2003).
* $p < 0.05$.

memory (Table 4). The finding that older age resulted in improved performance on tests of attention and processing speed over time is surprising. Closer inspection of studies that contributed data for this domain reveals that this finding is primarily because of one study (Dujardin et al., 2004), which observed fairly large improvement on two outcome measures (i.e., Stroop test color naming, Stroop interference test) in a sample of newly diagnosed PD patients. When this study was excluded from the analysis, no relationship was found between age and performance on tests of attention and psychomotor speed.

As shown in Table 5, educational level was associated with decline in all cognitive domains assessed. Specifically, fewer years of formal education were associated with a greater degree of decline in all domains. The influence of education was most pronounced on mental flexibility and reasoning ($p \leq .0001$), attention and processing speed ($p \leq .001$), and memory ($p \leq .01$). The duration of disease was not related to change in any of the cognitive domains assessed (data not shown).

Table 6 displays the results regarding the effects of the length of follow-up period on the magnitude of decline across

Table 4. The influence of age on effect sizes

Domain	Age (years)	k	Sample size Baseline	Sample size Follow-up	d	95%CI	Qw	QB
Global cognitive ability	<64	6	158	158	.24	−.01–.48	10.18	6.73*
	≥64	9	379	331	.54	.39–0.69	35.47†	
Verbal ability	<64	5	123	123	.15	−.16–.46	4.47	1.20
	≥64	3	116	106	−.02	−.52–.47	2.38	
Memory	<64	7	196	196	.18	−.02–.38	8.58	5.77*
	≥64	5	151	145	.49	.21–0.76	30.59†	
Verbal fluency	<64	6	146	154	.09	−.15–.33	10.98	.01
	≥64	5	170	164	.10	−.10–.29	12.19*	
Mental flexibility/Reasoning	<64	8	206	206	.20	−.01–.42	19.94*	1.14
	≥64	3	117	109	.05	−.45–.55	6.02	
Attention and processing speed	<64	5	120	120	.19	−.07–.45	10.37*	14.73†
	≥64	3	63	63	−.39	−.91–.13	11.69*	
Visuoconstructive skills	<64	3	102	101	.39	−.13–.92	4.08	.65
	≥64	4	126	116	.26	−.10–.62	1.51	

k = number of studies; d = mean weighted effect size; CI = confidence interval; Qw = homogeneity within a class of studies examining the same domain (df = k-1); QB = homogeneity between categories (df = 1).
* $p < .05$; † $p < .0001$.

Table 5. The influence of education on effect sizes

Domain	Education (years)	k	Sample size Baseline	Sample size Follow-up	d	95%CI	Qw	QB
Global cognitive ability	<10.9	4	116	116	.65	.29–1.01	4.32	5.94*
	≥10.9	4	169	169	.28	−.05–.60	8.36*	
Verbal ability	<10.9	2	89	79	.34	−1.40–2.08	.60	6.42*
	≥10.9	5	142	142	−.09	−.37–.19	.70	
Memory	<10.9	5	165	159	.56	.31–.80	16.80*	7.87**
	≥10.9	4	112	112	.15	−.21–.51	7.46	
Verbal fluency	<10.9	4	114	108	.39	.04–.74	2.99	3.85*
	≥10.9	4	132	132	.12	−.13–.37	1.89	
Mental flexibility/Reasoning	<10.9	4	155	147	.47	.13–.81	5.74	13.82****
	≥10.9	5	142	142	−.08	−.36–.20	.70	
Attention and processing speed	<10.9	2	61	61	.51	−1.13–2.15	.12	10.39***
	≥10.9	4	70	70	−.07	−.46–.33	.30	
Visuoperceptual functions	<10.9	3	93	84	.30	−.22–.82	1.46	5.91*
	≥10.9	3	63	62	−.14	−.72–.44	.57	
Visuoconstructive skills	<10.9	3	99	89	.56	−.01–1.13	.58	5.54*
	≥10.9	3	121	120	.16	−.31–.63	.11	

k = number of studies; d = mean weighted effect size; CI = confidence interval; Qw = homogeneity within a class of studies examining the same domain ($df = k − 1$); Q_B = homogeneity between categories ($df = 1$).
* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; **** $p \leq .0001$.

cognitive domains. Studies with follow-up intervals longer than 2 years yielded larger effect sizes for global cognitive ability ($p < .05$). The influence of age and the length of test-retest interval on visuoperceptual functions could not be assessed due to insufficient number of studies.

DISCUSSION

The results of the present meta-analysis indicate that non-demented PD patients show a relatively small decline in cognitive functions during an average follow-up period of

Table 6. The influence of the length of follow-up on effect sizes

Domain	Follow-up (months)	k	Sample size Baseline	Sample size Follow-up	d	95%CI	Qw	QB
Global cognitive ability	<24	9	260	260	.25	0.08–0.41	13.23	4.76*
	≥24	10	490	442	.45	0.32–0.57	58.44†	
Verbal ability	<24	6	120	110	.05	−0.25–.35	1.53	0.02
	≥24	2	119	119	.08	−1.31–1.47	6.50*	
Memory	<24	7	140	134	.32	.09–.54	28.37†	.03
	≥24	5	207	207	.29	.06–.53	16.55*	
Verbal fluency	<24	8	149	143	.02	−.18–.22	8.90	1.07
	≥24	4	177	177	.13	−.10–.37	14.65*	
Mental flexibility/Reasoning	<24	9	161	153	.01	−.22–.24	9.80	2.81
	≥24	3	172	172	.24	−.17–.66	15.69†	
Attention and processing speed	<24	6	88	88	−.06	−.34–.22	.70	.11
	≥24	3	105	105	−.01	−.42–.40	36.19†	
Visuoconstructive skills	<24	4	98	87	.27	−.15–.70	1.42	.21
	≥24	3	130	130	.35	−.10–.80	4.61	

k = number of studies; d = mean weighted effect size; CI = confidence interval; Qw = homogeneity within a class of studies examining the same domain ($df = k-1$); Q_B = homogeneity between categories ($df = 1$).
* $p < .05$; † $p < .001$.

2.5 years. Specifically, the magnitude of decline in the various cognitive domains ranged from 0 to 0.4 standard deviations. Among the cognitive domains analyzed, global cognitive ability, visuoconstructive skills, and memory showed statistically significant declines. In all domains the homogeneity between studies was acceptable, which suggests that the reported effects can be considered reliable.

Our findings suggest that there are subtle differences in the extent of decline across the various cognitive domains. Most prominent changes over time were detected on global dementia screening measures (e.g., MMSE), which summarize multiple aspects of cognitive function. One would therefore expect to find corresponding changes in several other cognitive domains. This was indeed the case for memory and visuoconstructive skills. In contrast, changes in the remaining domains were negligible. However, one should be cautious when comparing effect sizes. For example, effect sizes found for global cognitive ability were based on the sample of 702 patients, whereas effects for all other domains were based on considerably fewer subjects. Furthermore, it is likely that some heterogeneity exists with regard to the individual trajectories of cognitive decline. In some patients, for example, executive function may be the first domain to show decline, whereas in others memory may deteriorate first. Consequently, this results in smaller effect sizes for specific cognitive domains, and more pronounced effects on global cognitive measures because these latter tests are likely to detect changes that occur in any of the specific domains.

In light of the generally held clinical view that PD is accompanied by progressive cognitive impairment, our finding of a minimal decline in a number of cognitive domains is puzzling. Perhaps the most remarkable result of our meta-analysis is the absence of clear evidence for deterioration in executive function, whereas it is widely believed that this type of decline occurs already in the early stages of the disease (Levin & Katzen, 2005). Executive dysfunction should therefore have been reflected in relatively large effect sizes for verbal fluency, mental flexibility and reasoning, and attention and processing speed. One possible explanation for this lack of decline in executive abilities is that it occurs in the very early stage and then tapers off during the mid stages of the disease. The patients in our analysis, having a mean disease duration of 7.7 years, were in a mid stage. However, comparison of baseline scores on tests of executive functions with normative data (Strauss et al., 2006) revealed that PD patients in the present review exhibited only moderate decrements (mean z-score $= -0.6$), whereas impairments of attention and processing speed were much more prominent (z-score $= -1.1$). Effects of medication might in part account for this finding, since there is some evidence that levodopa treatment may improve executive functions of PD patients (Lange et al., 1992).

Another explanation for the present results could be a selection bias. Loss to follow-up was considerable in some of the studies. In general, the oldest and cognitively most compromised patients are the ones to drop out, causing

reductions of effect sizes in this type of research (Levin et al., 2000). This is underscored by the fact that only four studies (Bayles et al., 1996; Ebmeier et al., 1990; Stern et al., 1998; Tachibana et al., 1995) comprised patients with evidence of dementia at follow-up evaluation. A separate analysis of these studies revealed a mean effect size of 0.46 for global cognitive ability, whereas studies with non-demented patients yielded a mean effect size of 0.38. Moreover, only one study (Stern et al., 1998) including demented PD patients utilized a comprehensive neuropsychological test battery. Individual effect sizes generated from that study were in the small range ($g = .1–.4$) for visuoperceptual functions, naming ability, abstract reasoning and verbal fluency, whereas moderate to large declines ($g = 0.5–1.5$) were observed on constructive skills, global cognitive functioning and various aspects of memory. Thus, it seems likely that the absence of dementia in the majority of samples used in the present meta-analysis has led to smaller effect size estimates for some cognitive domains. Other meta-analyses that cross-sectionally compared neuropsychological test performance of PD patients and healthy control subjects have repeatedly demonstrated that impairments in the PD group are particularly robust when there is evidence of dementia (Henry & Crawford, 2004; Zakzanis & Freedman, 1999).

There are some methodological issues that might have affected our findings. This meta-analysis included longitudinal studies that used neuropsychological tests as dependent variables. An important issue related to repeated neuropsychological testing is practice or retest effects. With repeated exposure to the same task, novelty of testing is likely to diminish. Consequently, implicit knowledge of test demands and a reduction of potential anxiety associated with examination procedure may enhance the subsequent test performance. These factors might have obscured real decrements in cognitive functions. Observations from non-clinical populations (Basso et al., 1999; Dikmen et al., 1999) suggest that measures with a problem-solving component or measures requiring generation of a strategy (e.g. Wisconsin Card Sorting Test) tend to be particularly susceptible to practice effects. Thus, it is possible that the absence of change in some aspects of executive function may be an artifact of practice effects compensating real decline. Methods to control or reduce practice effects are the use of control groups, use of alternate forms of tests, and increase of test-retest intervals (McCaffrey & Westervelt, 1995).

In the present meta-analysis, only seven studies (Aarsland et al., 2004; Azuma et al., 2003; Bayles et al., 1996; Firbank et al., 2005; Hayashi et al., 1996; Katsarou et al., 1998; Serrano-Dueñas & Bleda, 2005) utilized a healthy control group. This was too small a number to allow comparison of cognitive decline in PD with normal cognitive aging (see also remarks below). The effect size for global cognitive ability obtained from studies that included a control group ($d = .45$) was similar to that for studies without control groups ($d = .38$). Furthermore, only three studies (Schmand et al., 2000; Smeding et al., 2006; Stebbins et al., 2000)

used alternate forms of some tests. With regard to the length of test-retest interval, 40% of the studies conducted follow-up assessment at rather short period of 12 months or less after baseline. Thus, few investigators have attempted to minimize the impact of practice effects. Future research in this field should be conducted more carefully and take the potentially confounding effects associated with repeated testing into account.

Another issue concerns the statistical model applied in the present analysis. We used the random-effects meta-analytical model rather than the more commonly employed fixed-effects model to obtain estimates of the mean effect sizes. Random-effects models are based on the assumption that in addition to sampling error, there is a true heterogeneity among the estimated population values as another source of variability that contributes to the observed differences in effect sizes across samples (Raudenbush, 1994). In comparison to fixed-effects models, random-effects analyses yield wider confidence intervals around the average effect size, thereby reducing the likelihood of committing a type I error. Random-effects analyses are therefore considered more conservative than fixed-effects models (Cohn & Becker, 2003). Because we anticipated heterogeneous results across studies, we reasoned that random-effects model would be more appropriate to calculate the pooled effect sizes. The advantage of random-effects analyses, however, is that they yield more generalizable parameter estimates, which extend beyond the studies included in the review (Cohn & Becker, 2003).

In order to better understand the relevance of the current findings it is important to consider these in relation to the cognitive changes in normal ageing. In the present review, we could not analyze reliably cognitive decline in healthy elderly subjects because only a few studies used a control group. A recent systematic review of the literature suggests that there is almost universal cognitive decline in the general elderly population, although the exact magnitude of this decline could not been determined (Park et al., 2003). However, studies on cognitive ageing in healthy subjects of comparable age to our PD sample and over a comparable follow-up interval, in general, suggest stability of cognitive performance. For example, a population-based study of elderly subjects using the MMSE as outcome measure showed no cognitive decline in younger age groups (65–69 years) three and five years after the initial assessment (Jacqmin-Gadda et al., 1997). One of the studies (Azuma et al., 2003) reviewed in the present analysis included a control group comprising subjects with comparable mean age (68 years), who were re-tested after approximately the same period of time (26 months) as the patients in this review. Performance on only one measure deteriorated significantly, whereas the scores on the remaining tests remained the same or improved. Thus, decline of cognitive functions in non-demented PD patients, albeit small, appears to be more pronounced than in healthy elderly subjects. The extent of cognitive changes ($d \leq .40$) in non-demented PD patients, however, if present, may be difficult to detect with small samples or with only clinical observation.

Whereas PD patients generally seem to show small changes in cognitive function over time, it is possible that certain sub-groups of patients exhibit more profound and obvious deterioration. It has, for example, been suggested that later onset of disease (Reid, 1992) and the presence of depressive symptoms at baseline (Starkstein et al., 1992) are associated with greater cognitive decline. Although the moderator analyses in the present review should be interpreted cautiously because of a small number of contributing studies, there are some interesting findings regarding the influence of demographic variables and study characteristics. Older age was found to be associated with greater degree of decline in global cognitive ability and memory. Although the effect of age was relatively small, it is consistent with previous studies examining the relation between demographic features and cognitive changes in PD (Palazzini et al., 1995; Portin & Rinne, 1986). Furthermore, our results indicate that education significantly influences changes in cognitive performance of PD patients over time. This suggests that high educational attainment may exert a protective effect on cognitive decline in PD, which is consistent with the cognitive reserve hypothesis (Stern, 2002). It has been reported before that low educational attainment is an important risk factor for the development of dementia in PD (Glatt et al., 1996). The present findings provide further evidence of the influence of education on decline in a number of major cognitive domains in PD. Our results show that duration of disease was not associated with the magnitude of cognitive changes.

Effect sizes for global cognitive ability were found to be greater in studies that employed longer follow-up intervals. This finding is consistent with the clinical experience that the progression of PD is rather slow, and corroborates the validity of the analysis. The apparent lack of effect of the follow-up interval on the remaining aspects of cognitive function is probably not a reliable finding, due to the small number of studies contributing to some analyses (e.g., verbal ability, attention and processing speed, and constructive skills).

This meta-analysis has several limitations, which are primarily related to the current state of the literature. First, although the number of patients in most comparisons was of reasonable size, a relatively small number of studies were included in analyses of some domains. Second, most studies have focused on only one or two cognitive domains. Consequently, the analyses of potential moderating variables were typically based on a small number of studies. In addition, only few studies have provided information about disease characteristics (e.g., severity of specific motor symptoms, medication, scores on depression rating scale), which prevented us from assessing the moderating effects of these variables. Third, when the same research group published multiple papers it was not always clear whether and to what degree there was overlap in the samples. This may have introduced bias in our analysis. However, in cases where there were indications that two or more studies were drawn from the same or highly overlapping samples, only the arti-

cle with the largest sample was included in the present review. Fourth, the classification of tests into cognitive domains is arbitrary to some degree. Fifth, about 70% of the longitudinal studies identified by the literature search did not meet the inclusion criteria for the present review. Closer inspection of demographic and study characteristics revealed that excluded studies had on average a longer follow-up interval (42 *versus* 29 months) and a shorter disease duration (5.1 *versus* 7.7 years) than studies included in the review. In light of our finding that the longer follow-up period was associated with a greater degree of cognitive decline, it could be argued that the exclusion of these studies might have resulted in some underestimation of the effect sizes. No differences were found between included and excluded studies with respect to age and educational level of patient samples. Finally, several studies have recruited patients from tertiary care clinics specialized in movement disorders. Therefore, it remains uncertain whether the present findings can be generalized to the PD population at large.

In conclusion, our quantitative review indicates that in non-demented PD patients, changes in cognitive functions over time are generally quite subtle. However, some methodological issues, such as the possibility of selection bias and the impact of practice effects, should be taken into consideration when interpreting the current findings. Clearly, more prospective studies are needed before firm conclusions can be drawn on the evolution of cognitive changes in PD. Future studies should include larger numbers of patients and employ comprehensive neuropsychological test batteries that evaluate various aspects of cognitive functioning. Several methodological issues should also be taken into account, for example, inclusion of a control group re-tested at the same interval and use of alternate test forms, in order to determine unequivocally the nature and magnitude of cognitive decline. Finally, future studies should attempt to measure the effects of potential moderator variables since they may be important in understanding the variability in the progression of cognitive changes in PD.

ACKNOWLEDGMENTS

REFERENCES[1]

*Aarsland, D., Andersen, K., Larsen, J.P., Perry, R., Wentzel-Larsen, T., Lolk, A., & Kragh-Sørensen, P. (2004). The rate of cognitive decline in Parkinson disease. *Archives of Neurology*, *61*, 1906–1911.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Association Press.

*Azuma, T., Cruz, R.F., Bayles, K.A., Tomoeda, C.K., & Montgomery, E.B. (2003). A longitudinal study of neuropsychological change in individuals with Parkinson's disease. *International Journal of Geriatric Psychiatry*, *18*, 1043–1049.

Basso, M.R., Bornstein, R.A., & Lang, J.M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *3*, 283–292.

*Bayles, K.A., Tomoeda, C.K., Wood, J.A., Montgomery, E.B., Cruz, R.F., Azuma, T., & McGeagh, A. (1996). Change in cognitive function in idiopathic Parkinson disease. *Archives of Neurology*, *53*, 1140–1146.

*Caparros-Lefebvre, D., Pecheux, N., Petit, V., Duhamel, A., & Petit, H. (1995). Which factors predict cognitive decline in Parkinson's disease? *Journal of Neurology, Neurosurgery, and Psychiatry*, *58*, 51–55.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohn, L.D. & Becker, B.J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, *3*, 243–253.

Della Sala, S., Di Lorenzo, G., Giordano, A., & Spinnler, H. (1986). Is there a specific visuo-spatial impairment in Parkinsonians? *Journal of Neurology, Neurosurgery, and Psychiatry*, *49*, 1258–1265.

Dikmen, S.S., Heaton, R.K., Grant, I., & Temkin, N.R. (1999). Test-retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, *5*, 346–356.

Dubois, B. & Pillon, B. (1997). Cognitive deficits in Parkinson's disease. *Journal of Neurology*, *244*, 2–8.

*Dujardin, K., Defebvre, L., Duhamel, A., Lecouffe, P., Rogelet, P., Steinling, M., & Destee, A. (2004). Cognitive and SPECT characteristics predict progression of Parkinson's disease in newly diagnosed patients. *Journal of Neurology*, *251*, 1383–1392.

*Ebmeier, K.P., Calder, S.A., Crawford, J.R., Stewart, L., Besson, J.A.O., & Mutch, W.J. (1990). Clinical features predicting dementia in idiopathic Parkinson's disease: A follow-up study. *Neurology*, *40*, 1222–1224.

*Firbank, M.J., Molloy, S., McKeith, I.G., Burn, D.J., & O'Brien, J.T. (2005). Longitudinal change in ^{99m}TcHMPAO cerebral perfusion SPECT in Parkinson's disease over one year. *Journal of Neurology, Neurosurgery, and Psychiatry*, *76*, 1448–1451.

Gib, W.R. & Lees, A.J. (1988). The relevance of the Lewy body to the pathogenesis of Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *51*, 745–752.

Glatt, S.L., Hubble, J.P., Lyons, K., Paolo, A., Tröster, A.I., Hassanein, R.E.S., & Koller, W.C. (1996). Risk factors for dementia in Parkinson's disease: Effect of education. *Neuroepidemiology*, *15*, 20–25.

*Growdon, J.H., Corkin, S., & Rosen, T.J. (1990). Distinctive aspects of cognitive dysfunction in Parkinson's disease. In M.B. Streifler, A.D. Korczyn, E. Melamed & M.B.H. Youdim (Eds.), Parkinson's disease: Anatomy, pathology, and therapy. *Advances in Neurology*, Vol. 53. New York: Raven Press.

*Hayashi, R., Hanyu, N., Kurashima, T., Tokutake, T., & Yanagisawa, N. (1996). Relationship between cognitive impairments, event-related potentials, and motor disability scores in patients with Parkinson's disease: 2-year follow-up study. *Journal of the Neurological Sciences*, *141*, 45–48.

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

---

[1]References marked with an asterisk were included in the meta-analysis.

Henry, J.D. & Crawford, J.R. (2004). Verbal fluency deficits in Parkinson's disease: A meta-analysis. *Journal of the International Neuropsychological Society*, *10*, 608–622.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

Hoehn, M.M. & Yahr, M.D. (1967). Parkinsonism: Onset, progression and mortality. *Neurology*, *17*, 427–442.

*Hovestadt, A. (1990). *Parkinson's disease, pulmonary function and muscle strength*. Rotterdam: Doctoral Dissertation.

*Hu, M.T.M., White, S.J., Chaudhuri, K.R., Morris, R.G., Bydder, G.M., & Brooks, D.J. (2001). Correlating rates of cerebral atrophy in Parkinson's disease with measures of cognitive decline. *Journal of Neural Transmission*, *108*, 571–580.

Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., & Dartigues, J.F. (1997). A 5-year longitudinal study of the Mini Mental State Examination in normal ageing. *American Journal of Epidemiology*, *6*, 498–506.

*Katsarou, Z., Bostantjopoulou, S., Alevriadou, A., Mentenopoulos, G., Avraam, X., & Kiosseoglou, G. (1998). A longitudinal study of visuospatial discrimination in parkinsonian patients. *Perceptual and Motor Skills*, *86*, 171–180.

Lange, K.W., Robbins, T.W., Marsden, C.D., James, M., Owen, A.M., & Paul, G.M. (1992). L-Dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Psychopharmacology*, *107*, 394–404.

Laupacis, A., Wells, G., Richardson, W.S., & Tugwell, P., for the Evidence-Based Medicine Working Group. (1994). Users' guides to the medical literature V: How to use an article about prognosis. *Journal of American Medical Association*, *3*, 234–237.

Levin, B.E., Llabre, M.M., Reisman, S., Weiner, W.J., Sanchez-Ramos, J., Singer, C., & Brown, M.C. (1991). Visuospatial impairment in Parkinson's disease. *Neurology*, *41*, 365–369.

Levin, B.E., Katzen, H.L., Klein, B., & Llabre, M.L. (2000). Cognitive decline affects subject attrition in longitudinal research. *Journal of Clinical and Experimental Neuropsychology*, *22*, 580–586.

Levin, B.E. & Katzen, H.L. (2005). Early cognitive changes and nondementing behavioral abnormalities in Parkinson's disease. *Advances in Neurology*, *96*, 84–94.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

McCaffrey, R.J. & Westervelt, H.J. (1995). Issues associated with repeated neuropsychological assessments. *Neuropsychology Review*, *3*, 203–221.

*Morrison, C.E., Borod, J.C., Perrine, K., Beric, A., Brin, M.F., Rezai, A., Kelly, P., Sterio, D., Germano, I., Weisz, D., & Olanow, C.W. (2004). Neuropsychological functioning following bilateral subthalamic nucleus stimulation in Parkinson's disease. *Archives of Clinical Neuropsychology*, *19*, 165–181.

Muslimovic, D., Post, B., Speelman, J.D., & Schmand, B. (2005). Cognitive profile of patients with newly diagnosed Parkinson's disease. *Neurology*, *65*, 1239–1245.

*Palazzini, E., Soliveri, P., Filippini, G., Fetoni, V., Zappacosta, B., Scigliano, G., Monza, D., Caraceni, T., & Girotti, F. (1995). Progression of motor and cognitive impairment in Parkinson's disease. *Journal of Neurology*, *242*, 535–540.

Park, H.L., O'Conell, E.O., & Thomson, R.G. (2003). A systematic review of cognitive decline in the general elderly population. *International Journal of Geriatric Psychiatry*, *18*, 1121–1134.

*Perrine, K., Dogali, M., Fazzini, E., Sterio, D., Kolodny, E., Eidelberg, D., Devinsky, O., & Beric, A. (1998). Cognitive functioning after pallidotomy for refractory Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *65*, 150–154.

Portin, R. & Rinne, U.K. (1986). Predictive factors for cognitive deterioration and dementia in Parkinson's disease. In M.D. Yahr & K.J. Bergmann (Ed.), *Advances in Neurology*, Vol. 45. New York: Raven Press.

*Ramirez-Ruiz, B., Marti, M.J., Tolosa, E., Bartres-Faz, D., Summerfield, C., Salgado-Pineda, P., Gomez-Anson, B., & Junque, C. (2005). Longitudinal evaluation of cerebral morphological changes in Parkinson's disease with and without dementia. *Journal of Neurology*, *11*, 1345–1352.

Raudenbush, S.W. (1994). Random effects model. In H. Cooper & L.V. Hedges (Ed.), *The handbook of research synthesis*. New York: Russell Sage Foundation.

Reid, W.G.J. (1992). The evolution of dementia in idiopathic Parkinson's disease: Neuropsychological and clinical evidence in support of subtypes. *International Psychogeriatrics*, *4*, 147–160.

Rosenberg, M.S., Adams, D.C., & Gurevitch, J. (2000). *MetaWin: Statistical software for meta-analysis*. Version 2. Sunderland, MA: Sinauer Associates.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

*Schmand, B., de Bie, R.M.A., Koning-Haanstra, M., de Smet, J.S., Speelman, J.D., & van Zomeren, A.H., for the Netherlands Pallidotomy Study (NEPAS) group. (2000). Unilateral pallidotomy in PD. A controlled study of cognitive and behavioural effects. *Neurology*, *54*, 1058–1064.

*Serrano-Dueñas, M. & Bleda, M.J. (2005). Delirium in Parkinson's disease patients. A five-year follow-up study. *Parkinsonism and Related Disorders*, *11*, 387–392.

*Smeding, H.M.M., Speelman, J.D., Koning-Haanstra, M., Schuurman, P.R., Nijssen, P., van Laar, T., & Schmand, B. (2006). Neuropsychological effects of bilateral STN stimulation in Parkinson disease. A controlled study. *Neurology*, *66*, 1830–1836.

*Soliveri, P., Monza, D., Paridi, D., Carella, F., Genitrini, S., Testa, D., & Girotti, F. (2000). Neuropsychological follow-up in patients with Parkinson's disease, striatonigral degeneration-type multi-system atrophy, and progressive supranuclear palsy. *Journal of Neurology, Neurosurgery, and Psychiatry*, *69*, 313–318.

*Starkstein, S.E., Bolduc, P.L., Mayberg, H.S., Preziosi, T.J., & Robinson, R.G. (1990). Cognitive impairment and depression in Parkinson's disease: A follow-up study. *Journal of Neurology, Neurosurgery, and Psychiatry*, *53*, 597–602.

*Starkstein, S.E., Mayberg, H.S., Leiguarda, R., Preziosi, T.J., & Robinson, R.G. (1992). A prospective longitudinal study of depression, cognitive decline, and physical impairments in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*, 377–382.

*Stebbins, G.T., Gabrieli, J.D.E., Shannon, K.M., Penn, R.D., & Goetz C.G. (2000). Impaired frontostriatal cognitive functioning following posteroventral pallidotomy in advanced Parkinson's disease. *Brain and Cognition*, *42*, 348–363.

*Stern, Y., Tang, M.X., Jacobs, D.M., Sano, M., Marder, K., Bell, K., Dooneief, G., Schofield, P., & Cote, L. (1998). Prospective comparative study of the evolution of probable Alzheimer's

disease and Parkinson's disease dementia. *Journal of the International Neuropsychological Society*, 4, 279–284.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 8, 448–460.

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

*Tachibana, H., Tomino, Y., Kawabata, K., Sugita, M., & Fukuchi, M. (1995). Twelve-month follow-up study of regional cerebral blood flow in Parkinson's disease. *Dementia*, 6, 89–93.

Zakzanis, K.K. & Freedman, M. (1999). A neuropsychological comparison of demented and non-demented patients with Parkinson's disease. *Applied Neuropsychology*, 3, 129–146.

Appendix A

Neuropsychological tests and corresponding cognitive domains

Domain	Test	k	%
Global cognitive ability	MMSE	12	48
	Mattis Dementia Rating Scale (MDRS)	2	8
	BDRS Information-Memory-Concentration	2	8
	CAMCOG	1	4
	Modified MMSE	1	4
	Mental Status Questionnaire	1	4
	Short Portable Mental Status Questionnaire	1	4
	Short Test of Mental Status	1	4
Verbal ability	Boston Naming Test	5	20
	WAIS Vocabulary	3	12
	WAIS Comprehension	1	4
	Peabody Picture Vocabulary Test	1	4
	Confrontation naming	1	4
	Verbal repetition	1	4
Memory	Auditory Verbal Learning Test (AVLT) (immediate recall = 4, delayed recall = 3, recognition = 3)	4	16
	Randt Memory Test Story recall (RMT) (immediate recall = 3, delayed recall = 1)	3	12
	California Verbal Learning Test (CVLT) (immediate recall = 1)	1	4
	Selective Reminding Test (SRT) (immediate recall = 1, delayed recall =1, recognition = 1)	1	4
	Hopkins Verbal Learning Test (HVLT-R) (immediate recall = 1, delayed recall = 1, recognition = 1)	1	4
	Grober & Buschke Test (immediate recall = 1, delayed recall = 1)	1	4
	WMS Logical Memory (immediate recall = 1)	1	4
	RBMT Logical Memory (immediate recall =1, delayed recall = 1)	1	4
	WMS Paired Associative Learning Test (delayed recall = 1)	1	4
	RMT Picture recognition test (immediate = 1, delayed = 1)	2	8
	Warrington Recognition Memory Test Faces (immediate = 1)	1	4
	Benton Visual Retention Test (Form D) (immediate = 1)	1	4
Verbal fluency	Category fluency	12	48
	COWAT letter fluency	10	40
Mental flexibility/Reasoning	(Modified) Wisconsin Card Sorting Test	6	24
	WAIS Similarities	4	16
	Raven Progressive Matrices	2	8
	Odd Man Out Test	2	8
	MDRS Concept formation subscale	1	4
	Word relations judgment	1	4
Attention and processing speed	Digit span forward & backward	4	16
	Stroop Test Part C (interference)	4	16
	Trail Making Test B	3	12
	Digit Ordering Test	1	4
	PASAT	1	4

Appendix A *Continued*

Domain	Test	k	%
Attention and processing speed (*continued*)	Brief Test of Attention	1	4
	Zazzo Test of Attention	1	4
	MDRS Attention subscale	1	4
	Stroop test A, B	4	16
	Trail Making Test A	3	12
	Symbol Digit Modalities Test	2	8
Visuoperceptual functions	Judgment of Line Orientation	4	16
	Hooper Visual Organization Test	1	4
	Benton Visual Retention Test (multiple choice)	1	4
	Overlapping Figures Test	1	4
	Visual Form Discrimination Test	1	4
	Test of Direction Sense	1	4
	GIT Spatial Test	1	4
Visuoconstructive skills	WAIS Block Design	5	20
	WAIS Object Assembly	2	8
	Clock Drawing Test	1	4
	Rosen Drawing Test	1	4
	Mattis DRS Construction subscale	1	4

k = number of studies; % = percent of articles in the meta-analysis that included the test to quantify change in cognitive functioning over time; MMSE = Mini Mental State Examination; BDRS = Blessed Dementia Rating Scale; CAMCOG = Cambridge Cognitive Examination; WAIS = Wechsler Adult Intelligence Scale; WMS = Wechsler Memory Scale; RBMT = Rivermead Behavioral Memory Test; COWAT = Controlled Oral Word Association Test; PASAT = Paced Auditory Serial Addition Test; GIT = Groninger Intelligence Test.