# Executive Dysfunction in Parkinson's Disease: A Meta-Analysis on the Wisconsin Card Sorting Test Literature

Florian Lange[1,2], Carolin Brückner[1], Aylin Knebel[2], Caroline Seer[2,3], & Bruno Kopp[2]

[1]Behavioral Engineering Research Group, KU Leuven, Leuven, Belgium

[2]Department of Neurology, Hannover Medical School, Hannover, Germany

[3]Movement Control & Neuroplasticity Research Group, KU Leuven, Leuven, Belgium

*Correspondence:*

*Florian Lange*

*BEE – Behavioral Engineering Research Group*

*KU Leuven*

*Naamsestraat 69*

*3000 Leuven*

*Belgium*

*florian.lange@kuleuven.be*

*Note*: This preprint may slightly differ from the article published in *Neuroscience & Biobehavioral Reviews* (https://doi.org/10.1016/j.neubiorev.2018.06.014).

## Abstract

Executive dysfunctions are a frequently described non-motor symptom in patients with Parkinson's disease (PD). However, the nature, extent, variability, and determinants of executive dysfunctions in PD are still poorly understood. To improve the characterization of executive dysfunctions in PD, we conducted a meta-analysis of the studies administering the Wisconsin Card Sorting Test (WCST) to patients with PD and healthy controls. We included $k = 161$ studies, which allowed us to precisely estimate the size of PD-related WCST deficits and to run powerful tests for potential moderators of these deficits. We found robust WCST deficits in PD, which were medium-to-large in size. These deficits were most pronounced in patients tested after withdrawal from dopaminergic medication and in samples characterized by severe motor impairment and long disease duration. Substantial WCST impairment was also detected in non-demented, non-depressed, and never-medicated patients with PD as well as after conservatively correcting for publication bias. Based on these findings, impaired WCST performance can be considered as a major hallmark of executive dysfunction in PD.



*Keywords*: Parkinson's disease; executive dysfunction; cognitive flexibility; Wisconsin Card Sorting Test; Meta-analysis

## 1 Introduction

In addition to characteristic motor symptoms such as bradykinesia, rigidity, and tremor, many patients with idiopathic Parkinson's disease (PD) show deficits in cognitive functioning (Zgaljardic, Borod, Foldi, & Mattis, 2003). While cognitive impairment in PD appears to be heterogeneous in nature (Kehagia, Barker & Robbins, 2013; Miller, Neargarder, Risi, & Cronin-Golomb, 2013; Robbins & Cools, 2014; Seer, Lange, Georgiev, Jahanshahi, & Kopp, 2016), the domain of executive functioning has received particular attention over the past decades. The term 'executive functioning' refers to a set of higher-order cognitive processes that enable goal-directed behavior and adjustments to novel situations by exerting top-down influence on lower-level cognitive processes (Friedman & Miyake, 2017). When executive functions are impaired, behavior becomes uncoordinated and disinhibited, rendering the individual inflexible and susceptible to distraction (Elliot, 2003). It is thus not surprising that executive dysfunctions are related to reduced quality of life in patients with PD and their caregivers (Kudlicka, Clare, & Hindle, 2014). In addition, the presence of executive dysfunctions in patients with PD has been shown to predict progression to Parkinson's disease dementia (PDD) (Janvin, Aarsland, & Larsen, 2005; Levy et al., 2002; Mahieux et al., 1998; Williams-Gray, Foltynie, Brayne, Robbins, & Barker, 2007; Woods & Tröster, 2003). Against this background, understanding the nature and extent of executive dysfunctions in PD is of critical importance.

Executive dysfunctions in PD have most frequently been examined by means of standardized neuropsychological tests. One of the most popular instruments in this literature is the Wisconsin Card Sorting Test (WCST; Berg, 1948; Grant and Berg, 1948; Heaton, Chelune, Talley, Kay, & Curtiss, 1993; Nelson, 1976).

The WCST requires participants to sort cards in accordance with one of three task rules (color, shape, number). The currently prevailing task rule (or sorting category) is not explicitly

revealed to participants. Participants have to test rules and to evaluate the examiner's feedback in order to identify the correct rule. After a predefined number of consecutive correct sorts by this rule, the category is considered to be completed and the valid task rule changes (see Figure 1). Card sorts according to the previously correct rule will then result in negative feedback. Participants are required to flexibly respond to this feedback by shifting to a new rule. Once the new rule has been identified, participants have to keep sorting by this rule until the next category is completed. The number of completed categories given a constant number of trials is frequently used as a measure of overall WCST performance. Moreover, a large number of additional performance measures have been proposed as indicators of more specific cognitive processes (Heaton et al., 1993; see Figure 1). Most prominently, deficits in cognitive flexibility (one of the core executive functions, Miyake et al., 2000) are commonly thought to be reflected in the number of perseverative errors committed on the WCST (Lange, Kröger, et al., 2016). A perseverative error is scored when a participant keeps sorting by a particular WCST rule although the experimenter's feedback has signaled that this rule is no longer valid.

As early as 1983, Lees and Smith reported that newly diagnosed patients with PD completed significantly less WCST categories and committed significantly more perseverative errors than healthy matched control participants (HC). These and related findings have been taken to support a link between the PD-specific dysfunction of the basal ganglia and deficient executive functioning. In the decades to follow, this hypothesis has received additional support from neuroimaging studies (Christopher & Strafella, 2013; Monchi, Hanganu, & Bellec, 2016). Contemporary models of basal ganglia contributions to executive functioning (Frank, Loughry, & O'Reilly, 2001; Hazy, Frank, & O'Reilly, 2007; Herd et al., 2014) have been critically informed by the evidence for WCST deficits in patients with PD.

The neuropsychological research design applied by Lees and Smith (1983) has been replicated more than a hundred times in various samples of patients with PD. Despite this wealth of research, the literature on WCST deficits in patients with PD has remained largely unintegrated. An early review (Lees, 1989) of six studies suggested that PD is consistently associated with WCST performance deficits, but that the nature and extent of these deficits may differ across studies. More than 20 years later, a meta-analysis by Kudlicka, Clare, and Hindle (2011) reported medium-to-large effect sizes ($g = 0.43 – g = 0.69$) with regard to the difference in WCST performance between patients with PD and HC. This meta-analysis also revealed the presence of substantial between-study heterogeneity in effect sizes. Medium-sized average effects resulted from a combination of some studies with very large group differences (e.g., Tomer, Fisher, Giladi, & Aharon-Peretz, 2002) and other studies with small PD-related WCST deficits (e.g., Cooper, Sager, Jordan, Harvey, & Sullivan, 1991). As Kudlicka and colleagues (2011) only included eight WCST studies in their meta-analysis, they were not able to identify the factors that account for this variability in effect sizes.

The small number of studies meta-analyzed by Kudlicka and colleagues (2011) likely resulted from the strict inclusion criteria applied in that meta-analysis. To be included, studies had to be explicitly based on a neuropsychological perspective and to directly state that the main goal of the study was “to investigate executive impairment in PD” (Kudlicka et al., 2011, p. 2307). In addition to limiting the precision with which effect sizes can be estimated and the possibilities for identifying moderating factors, these inclusion criteria might be associated with another methodological problem. Requiring studies to explicitly focus on executive impairment in PD might exclude some of those studies that did not restrict their exploration of potential PD-related alterations to the domain of executive functioning, but also administered other potentially interesting measures. Depending on the perceived conclusiveness and significance of the results,

the authors of such studies might decide to focus their report on one set of measures or another. As a corollary, studies that do not find conclusive evidence for executive impairment in PD might be less likely to be reported in an article with the explicit goal of investigating executive impairment in PD and hence less likely to be included in the meta-analysis by Kudlicka and colleagues (2011). This type of publication bias might lead to a substantial overestimation of PD-related deficits on the WCST. As the small number of studies included by Kudlicka and colleagues (2011) does not allow for powerful tests for publication bias, conducting a new, more inclusive meta-analysis is the most promising way to arrive at more reliable evidence regarding potential WCST performance deficits in patients with PD.

Here, we present a comprehensive meta-analytic overview of the studies comparing WCST performance between patients with PD and HC. Our search strategy and inclusion criteria led to the inclusion of effect sizes from more than 150 studies. The richness of this data set allowed us to pursue four main study goals with high statistical power. First, we aimed to precisely determine the extent and variability of WCST performance deficits in patients with PD. Second, we compared PD-related WCST deficits across different WCST measures to examine whether some aspects of WCST performance are more affected than others. Third, we investigated whether the size of WCST performance deficits in patients with PD is moderated by characteristics of the examined sample. By this means, we were able to test whether between-study variability in WCST deficits can be accounted for by differences in the severity of motor impairment, disease duration, and medication status, among others. Similarly, it was possible to determine whether study quality (i.e., the degree to which patients and HC were matched with regard to sociodemographic variables) affects the magnitude of reported WCST deficit in patients with PD. Fourth, we estimated the extent to which our results are affected by publication bias and took a series of measures to adjust for any potential biases. In combination, these analyses

allowed investigating if PD is accompanied by substantial WCST performance deficits, how large these deficits are, and under which circumstances they are most pronounced.

## 2 Methods

### 2.1 Search strategy

A systematic literature review was conducted in 2015 and updated in May 2017. We searched for records including the term "Parkinson" in combination with any of the three following keywords: "card sorting", "WCST", "MCST". Google Scholar (12,425), PubMed (113), PsycNet (439), and Web of Science (184) yielded a total of 13,211 hits for these combinations of search terms (Figure 2).

We screened the titles and abstracts of these records to exclude studies that did not report any original WCST data obtained from patients with PD. Each record was screened by at least one author (CB or AK). When this author was not sure whether a record can be excluded, she discussed the case with a second author (FL). We accessed the full text of those records that we did not exclude based on this criterion. Where full texts were not accessible online or via local university libraries, we attempted to contact the original authors. In total, we accessed 616 full texts.

In a next step, we excluded 455 of these papers because they did not fulfill all of the following inclusion criteria.

1) A standard version of the WCST had to be administered to a sample of patients with PD as well as to a sample of healthy control participants (HCs). Non-standard WCST versions (e.g., computerized paradigms for the assessment of response times) were excluded when their outcome measures did not directly relate to the standard WCST measures distinguished in this meta-analysis.

2) The article had to report data for at least one WCST measure at a level of detail that allows for the calculation of effect sizes. Articles were included when they provided means and standard deviations for patients with PD and HCs or the test statistic for the between-group comparison in WCST performance. We also included articles reporting descriptive data (median and range or median and interquartile range) that allow for estimating means and standard deviations according to the procedure described by Wan, Wang, Liu, and Tong (2014).

3) The WCST data reported in the paper had to be unique. When the same (or partially overlapping) data were reported in multiple papers, we included that record which provided the most comprehensive WCST data (e.g., more outcome variables) or data from a larger sample of participants. When we considered multiple papers equally informative, we selected the record with the earliest publication date.

We explicitly included papers written in languages other than English if WCST data relevant for effect-size calculation were identifiable without ambiguity. We retained 161 records that fulfilled the criteria listed above. Each record was screened by at least one of the authors and a randomly selected subset ($n$ = 30) of the accessed full texts was screened independently by two of the authors (C.B. & F.L.) to determine inter-rater reliability (IRR) of the inclusion procedure. Both authors identified the same seven of these records as eligible for inclusion ($\kappa = 1.00$).

**2.2 WCST outcome measures**

We performed separate meta-analyses for those established measures of WCST performance that have been reported in at least 10 of the included studies (see Figure 1, for illustration of these measures). This criterion was set to guarantee a minimum of statistical power for all analyses and to prevent the number of analyses from being inflated by the inclusion of rarely used or idiosyncratic measures. Analyzed measures include: 1) the number of completed categories, 2) the number of perseverative errors, 3) the percentage of perseverative errors, 4) the

number (or percentage) of perseverative responses, 5) the number (or percentage) of non-perseverative errors, 6) the total number (or percentage) of errors, 7) the number of trials required to complete the first criterion, 8) the number of failures to maintain set, 9) the percentage of conceptual level responses, and 10) global scores of WCST performance. While we were able to distinguish between the number and percentage of perseverative errors, making the same distinction for other outcome measures was deemed impractical due to relatively small numbers of studies reporting percentage values for these measures. When studies did not report the total number (or percentage) of errors but the total number (or percentage) of correct responses, we used the latter measure and changed the sign of the extracted effect size. The outcome-measure category "global scores" includes diverse aggregate measures reported in the included studies.

We selected the two most frequently reported variables as principal outcome measures for additional in-depth analyses of WCST performance deficits in patients with PD. We observed that the vast majority of the included articles reported at least one measure of perseveration. To avoid redundancy and increase statistical power, we selected one measure of perseveration for each of these studies (cf. Demakis, 2003). This measure will be referred to as "perseverations" in the following. When multiple measures of perseveration were reported, we selected the measure for the perseveration variable according to the following hierarchy: number of perseverative errors, percentage of perseverative errors, number of perseverative responses, percentage of perseverative responses. Similarly, most of the included articles reported the number of WCST categories completed by patients with PD and HC. Hence, we selected the number of completed categories as the second principal outcome measure.

### 2.3 WCST Data Extraction and Effect Size Calculation

When articles reported means and standard deviations of WCST outcome variables for patients and HCs, we calculated the *t*-statistic for the between-group comparison as defined by

Welch's *t*-test. For studies that did not provide these but other descriptive statistics (i.e., group medians as well as either minima and maxima or interquartile range), means and standard deviations were estimated using the procedure described by Wan and colleagues (2014) and subsequently used to calculate the *t*-statistic. For studies that did not report sufficient descriptive data, but the *t*-statistic of the corresponding group comparison, we used this *t*-statistic as long as the direction of the outcome (PD-related improvement vs. deficit) was unambiguous. For a subset of 37 studies (the seven included studies used for determining the IRR of the inclusion procedure plus 30 additional randomly selected studies), effect-size relevant data for our two principal outcome measures was extracted and *t*-statistics were determined by two independent raters (C.B. & F.L.). The inter-rater Pearson correlation between *t*-statistics was $r = .95$ for the number of completed categories and $r = .93$ for perseverations.

Effect sizes (Cohen's *d*) and their 95% confidence intervals (CIs) were calculated from *t*-statistics using the SPSS syntaxes provided by Wuensch (2012). When the value of the *t*-statistic was 0 due to floor or ceiling effects (e.g., when both groups completed an average of six categories with a standard deviation of zero), this procedure does not allow estimating a confidence interval. In these cases, we replaced the *t*-statistic by 1, estimated the size of the CI, and centered it around 0. When an article did not report any of the data mentioned above but the *F*-value of a between-subjects ANOVA (with the difference being unambiguous in direction), Cohen's *d* was calculated using the procedure provided by Lenhard and Lenhard (2016). For studies reporting only the test statistic (*z*) of a Wilcoxon Signed Rank Test, *z* was divided by the square root of the sample size to obtain *r* which then was transformed to *d* (Field, 2013). When a study involved more than one group of patients with PD (e.g., tremor-dominant vs. akinetic and rigidity-dominant patients, Yu, Wu, Tai, Lin, & Hua, 2010), data were pooled across groups (unless the subgroups were divided according to one of our *a priori* defined moderator variables,

see below). This procedure resulted in effect sizes being extracted from a total of 180 samples of PD patients (see Table 1). Effect sizes were transformed such that more positive values indicate more pronounced deficits in patients with PD.

### 2.4 Basic Meta-Analysis

Mean effect sizes and confidence intervals for our two principal WCST outcomes as well as the other nine WCST measures were calculated using the random-effects model SPSS syntax provided by Field and Gillett (2010). A random-effects model was chosen because we assumed the true extent of PD-related WCST deficits to differ systematically between studies (e.g., as a function of the included moderator variables). Heterogeneity of effect sizes was examined using Cochran's $Q$ and the $I^2$ index (Higgins, Thompson, Deeks, & Altman, 2003). By comparing Cochran's $Q$ (estimated under fixed-effect assumptions) to a $\chi^2$ distribution, we tested whether heterogeneity among studies was significant. The $I^2$ index served as an estimate of between-study variability in true effect sizes, with $I^2$ values of about 25%, 50% and 75% indicating low, moderate, and high heterogeneity, respectively (Higgins et al., 2003). We also performed a meta-analysis on the difference between the effect sizes of our two principal outcome measures. To this end, we subtracted the effect size for perseverations from the effect size for categories for every study that reported data for both variables, and applied the above mentioned random-effects model syntax to the effect-size difference. By this means, it was possible to analyze whether one of our principal WCST outcome measures was significantly more affected than the other in patients with PD.

### 2.5 Moderator analysis

Our two principal WCST outcome measures were also used to investigate potential moderators of WCST performance deficits in patients with PD. Specifically, we tested whether effect sizes for the comparison between patients and HCs varied as a function of various sample

characteristics or indicators of study quality. We selected the following sample characteristics as potential moderators: 1) the mean age of patients in the PD group, 2) the proportion of female participants in the PD group, 3) the mean disease duration in the PD group, 4) the mean HY stage in the PD group, 5) the mean score on the motor scale of the Unified Parkinson's Disease Rating Scale (UPDRS), 6) the mean score of the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), 7) the medication status of patients during the time of neuropsychological examination, 8) the exclusion of patients with dementia in the PD group, 9) the exclusion of patients with depression in the PD group.

When only a range for patients' HY stages was provided (e.g., stage I – II), we used the mean between these stages (in this case, 1.5) as an estimate for the mean HY stage in the PD group unless the provided range was too large (i.e., larger than three stages) to render meaningful information. For studies reporting an HY stage range larger than three, we did not attempt to estimate mean HY stage and these studies were excluded from the analysis of this moderator. When studies provided HY or UPDRS values for both patients' ON (i.e., with dopaminergic medication) and OFF (i.e., without dopaminergic medication) state, we selected the measurement that corresponded to the medication status in which patients were examined with the WCST. With regard to the medication status, we distinguished between studies that included only patients who were examined ON medication and studies that included only patients who were examined OFF medication. Within the latter category, we additionally distinguished between unmedicated patients who had never received dopaminergic medication (de novo) and patients who had undergone a medication washout period prior to neuropsychological testing (withdrawal). A relatively large number of samples included both medicated and unmedicated patients (see Table 1) and these studies have been excluded from the analysis of this potential moderator. With regard to the presence of dementia, we distinguished between studies that

excluded patients with dementia and studies that did not exclude patients with dementia. A study was coded as excluding patients with dementia when it explicitly mentioned that none of the patients showed signs of dementia. In the large majority of these studies, it was not specified which criterion had been used to exclude patients with dementia. Most of the studies that did provide this information used an MMSE cut-off score of 24 to screen for dementia. To apply a consistent criterion across all studies, we also coded studies as excluding patients with dementia when no explicit exclusion statement was given, but when we could ascertain that all included patients scored higher than 24 on the MMSE. We note, however, that an MMSE score of 24 or lower is commonly considered to be neither necessary nor sufficient for a diagnosis of PDD (Dubois et al., 2007; Emre et al., 2007). Furthermore, we applied a rather conservative criterion to distinguish between studies that excluded depressed patients and studies that did not exclude depressed patients. In order for a study to be coded as excluding depressed patients, the study was required to (a) explicitly mention depression as an exclusion criterion and (b) report a smaller than medium difference ($d < 0.5$) between patients with PD and healthy controls on a depression rating scale. If patients' performance on the WCST is found to be impaired in these studies, it is rather unlikely that PD-related WCST impairment is secondary to depression. Originally, we planned to also evaluate the moderating role of neurosurgical procedures on PD-related WCST deficits. However, across all studies, only three studies (Ravizza & Ciranni, 2002; Smith & MacDowall, 2006a,b) reported having included small numbers of patients with PD who had undergone pallidal surgery (three patients in total) or deep-brain stimulation (one patient). We thus refrained from including this sample characteristic in our moderator analysis.

As indicators of study quality, we used three measures that reflect how well patients with PD and HCs had been matched. Specifically, we selected 1) the difference between the mean age in the PD group and the mean age in the control group, 2) the difference between the proportion

of female participants in the PD group and the proportion of female participants in the control group, and 3) the difference between the mean MMSE score in the PD group and the mean MMSE score in the control group.

We determined IRR for the extraction of moderator variables from the individual studies according to the same procedure as described for the extraction of effect sizes (see above). Inter-rater Pearson correlation coefficients were larger than .9 for seven of the nine continuous variables, .78 for the proportion of female participants in the patient sample and .23 for the gender proportion difference between the PD group and the HC group. The latter two values resulted from an isolated coding error made during data extraction from a single study, which we corrected before running the meta-analyses (corrected $r = 1.00$). IRR for the three categorical variables (medication status, depression and dementia) was $\kappa = 1.00$. To facilitate comparison between predictors, all continuous variables were z-transformed before we conducted the moderator analyses.

The relationship between these nine continuous and three categorical predictors and PD-HC group differences in WCST categories and perseverations was examined using separate weighted multiple regression analyses (Field & Gillett, 2010). In a subsequent step, we included all significant predictors in the same meta-regression model to determine which, if any, variable explains unique variance in the size of PD-related WCST performance deficits.

Note that we report results on an additional categorical moderator variable that might be related to study quality. During the review process, we were alerted of the possibility that data extracted from unpublished studies (which did not undergo peer-review) or from studies published in a language other than English (which are more difficult to screen for the relevant information) might be less reliable. As a consequence, we analyzed whether effect sizes and their heterogeneity differed between those studies and studies published in English journals by adding

"publication status" (0 = published and English, 1 = unpublished or non-English) to our moderator analyses.

### 2.6 Publication Bias Analysis

We took a series of measures to prevent, assess, and adjust for the possible influence of publication bias (i.e., the overrepresentation of studies showing statistically significant results due to their selective publication in scientific journals). First, we did not limit our search of relevant studies to the literature published in journals with peer-review, but also included theses and dissertations that are indexed in Google Scholar. Second, we ran follow-up robustness analyses including only non-significant effect sizes. By definition, it can be excluded that this sample of non-significant effect sizes is affected by publication bias. Hence, when mean effect sizes for PD-related WCST performance deficits are still significantly larger than zero in this subset of studies, it can be excluded that the evidence for these deficits purely results from the selective publication of significant results. Third, the Begg and Mazumdar's rank correlation test was calculated as implemented in the syntax by Field and Gillett (2010) to examine the relationship between effect sizes and their standard errors. A positive correlation between these two variables would indicate an overrepresentation of small studies with large effect sizes. Such a small-study effect can be the result of publication bias and it would likely contribute to an overestimation of the true effect size. In an attempt to adjust for possible relationships between sample size and effect size, we ran weighted linear regression analyses with effect sizes as outcome variable, the inverse of sample sizes as predictor variable, and sample sizes as weights (Peters, Sutton, Jones, Abrams, & Rushton, 2006). The model's intercept is interpreted as a tentative estimate of the effect size in a perfectly precise (i.e., infinitely large) study. Finally, we used the weight functions proposed by Vevea and Woods (2005) and implemented by Field and Gillett (2010) in SPSS and R to examine the degree to which mean effect sizes change under different selection bias models. The four

implemented models reflect the assumptions that 1) significant studies in reporting PD-related WCST deficits have a moderately increased chance of being published (moderate one-tailed selection), 2) significant studies in reporting PD-related WCST deficits have a severely increased chance of being published (severe one-tailed selection), 3) significant studies in either direction (PD-related WCST deficits or improvements) have a moderately increased chance of being published (moderate two-tailed selection), and 4) significant studies in either direction (PD-related WCST deficits or improvements) have a severely increased chance of being published (severe two-tailed selection). The degree to which effect sizes differ between the results of our random-effects meta-analyses and these selection-model analyses reflects the robustness of the effect-size estimates against the assumption that they have been produced by publication bias (Field & Gillett, 2010).

# 3 Results

## 3.1 WCST Deficits in Patients with PD

Patients with PD performed significantly worse than HC on all of the meta-analyzed WCST measures (see Table 2). PD-related WCST performance deficits were medium-to-large in size for most measures and ranged from $d = 0.29$ (failures to maintain set) to $d = 0.78$ (total number of errors). Due to the large number of included studies, we were able to estimate effect sizes with considerable precision as reflected in the narrow confidence intervals displayed in Table 2. Being based on more than 140 samples and involving over 7500 participants, the analyses of PD-related deficits on our two main outcome measures (categories, perseverations) were particularly powerful. While patients with PD showed substantial impairment on both of these measures, PD-related WCST deficits seem to be larger with regard to the number of completed categories, $d = 0.74$, 95% CI [0.67, 0.82], than with regard to the number of

committed perseverations, $d$ = 0.57, 95% CI [0.49, 0.63]. Note that the CIs surrounding the two effect sizes do not overlap, suggesting that the magnitude of PD-related WCST performance deficits differs significantly across measures. To test this idea more directly, we conducted a follow-up analysis involving those studies that allowed calculating effect sizes for both the number of completed categories and the number of committed perseverations. For each of these $k$ = 118 samples, we calculated the difference between the two corresponding effect sizes ($\Delta d$). A meta-analysis of effect-size differences revealed that PD-related deficits on the category measure were indeed significantly larger than on the perseveration measure, $\Delta d$ = 0.14, 95% CI [0.08, 0.20].

### 3.2 Publication Bias Analysis

Across all analyzed measures, the effect sizes extracted from individual studies were positively associated with their standard errors as indicated by Begg and Mazumdar's rank correlation test. Correlations were small-to-medium in size and reached statistical significance for five of the analyzed WCST variables (categories, perseverations, perseveration errors (n), non-perseverative errors, total errors). These results suggest that the effect sizes from our random-effects meta-analyses may be overestimated due to publication bias or another type of small-sample bias in the analyzed set of studies. However, we ran a number of additional robustness analyses suggesting that the influence of this kind of bias on our effect-size estimates is rather small (Table 3). First, when we repeated our analyses including only the studies that reported a non-significant difference between patients with PD and HC in their performance on the WCST, average effect sizes remained significantly larger than zero in all but one case (failures to maintain set). Second, when we regressed effect sizes on the inverse of the associated sample sizes, the obtained corrected effect-size estimates (i.e., the intercepts in the regression model) decreased only slightly in comparison to the effect-size estimates from our random-effects

analysis and remained significantly larger than zero in all but three cases (trials to criterion, failures to maintain set, global score). Third, application of the selection bias models proposed by Vevea and Woods (2005) showed that effect-size estimates decrease only marginally, even if one assumes a severe selection bias in favor of studies reporting significant results. Two results that stood out in the latter analysis were the large negative estimates for group differences with regard to failures to maintain set and conceptual level responses when severe one-tailed publication bias was assumed. These implausible figures seem to be due to the presence of some instances of small and non-significant performance improvements in patients with PD in the small set of studies reporting these WCST measures. Removing these studies with negative effect sizes renders the results of the selection model analysis for failures to maintain set and conceptual level responses comparable to the results for other WCST variables.

In sum, for some of the analyzed WCST measures that have not been reported in a large number of studies, our robustness analyses did not unequivocally support the presence of significant deficits in patients with PD. In contrast, the available data revealed robust PD-related deficits on more established WCST measures (e.g., categories, perseverations, total errors) that are very unlikely to result from publication bias.

### 3.3 Heterogeneity and Moderator Analyses

Effect-size heterogeneity ranged from negligible (non-perseverative errors, trials to criterion) to large (conceptual level responses) values, and was moderate (i.e., around $I^2 = 50\%$) for most of the analyzed WCST measures. These results indicate that the size of PD-related deficits on the WCST may vary as a function of sample characteristics or study quality. To address this possibility, we conducted a series of moderator analyses using our two principal WCST outcome measures (categories and perseverations). As can be seen from inspection of Table 4, PD-related WCST deficits with regard to perseverations were not significantly

moderated by sample characteristics (age, gender, disease duration, HY stage, UPDRS motor score, MMSE score, dementia status, depression status, medication status) or by indicators of matching quality (PD vs. HC differences in age, gender, and MMSE scores). Similarly, effect sizes did not differ as a function of publication status. Unpublished studies or studies published in a non-English language yielded effect sizes (categories: $d$ = 0.70, 95% CI [0.47, 0.93], $I^2$ = 44.48%, perseverations: $d$ = 0.49, 95% CI [0.22, 0.75], $I^2$ = 34.82%) that were similar to those reported in published English journal articles (categories: $d$ = 0.75, 95% CI [0.67, 0.83], $I^2$ = 58.21%, perseverations: $d$ = 0.57, 95% CI [0.50, 0.65], $I^2$ = 52.21%).

In contrast, longer disease duration, ß = .09, $t(120) = 2.13$, $p = .036$, and higher scores on the UPDRS motor scale, ß = .16, $t(69) = 3.08$, $p = .003$, predicted larger PD-related WCST deficits on our second main outcome measure (i.e., the number of completed categories). Deficits on the category measure also varied as a function of medication state, $\chi^2(2) = 8.24$, $p = .016$ (see Figure 3). Studies that exclusively included never medicated *de novo* patients found only small differences between PD patients and HC, $d$ = 0.35, 95% CI [0.20, 0.49], $I^2$ = 8.05%. Deficits were larger in patients who were tested on their usual dopaminergic medication, $d$ = 0.76, 95% CI [0.66, 0.86], $I^2$ = 56.81%, and largest in patients who were tested during withdrawal of their usual medication, $d$ = 1.13, 95% CI [0.60, 1.65], $I^2$ = 62.85%. When the three significant predictors of PD-related deficits in the number of completed WCST categories were entered simultaneously, only the UPDRS score, ß = .18, $t(41) = 2.64$, $p = .012$, but neither disease duration, ß = -.02, $t(41) = -0.38$, $p = .708$, nor medication status, $\chi^2(2) = 3.42$, $p = .181$, emerged as a significant predictor. To further characterize the relationship between motor impairment (as measured by the UPDRS) and WCST performance deficits in patients with PD, we calculated effect sizes separately for the four quartiles of studies distinguished according to the mean UPDRS score of the included patients (Figure 4). As can be seen from Figure 4, effect sizes ($d$ = 0.61, $d$ = 0.53, $d$ = 0.62) do

not vary substantially across the first three UPDRS quartiles (with mean UPDRS scores of $M$ = 12.63, $M$ = 17.12, $M$ = 21.11). However, in contrast to these first three UPDRS quartiles, WCST performance deficits in patients with PD were considerably increased ($d$ = 1.09) in the set of studies reporting average UPDRS scores in the highest quartile ($M$ = 29.22). Of note, PD-related WCST performance deficits were also substantial in the subset of studies excluding patients with dementia (categories: $k$ = 108, $d$ = 0.70, 95% CI [0.61, 0.79], $I^2$ = 56.27%, perseverations: $k$ = 113, $d$ = 0.55, 95% CI [0.47, 0.63], $I^2$ = 3.27%) as well as in the small number of studies fulfilling our conservative criteria for excluding patients with depression (categories: $k$ = 7, $d$ = 0.77, 95% CI [0.54, 1.01], $I^2$ = 0%, perseverations: $k$ = 7, $d$ = 0.83, 95% CI [0.55, 1.11], $I^2$ = 26.67%).

## 4 Discussion

Our meta-analysis of WCST performance alterations in patients with PD revealed three key findings. First, in contrast to healthy controls, patients with PD showed significant impairment across all examined WCST measures. These deficits were medium-to-large in size and remained robust even when we conservatively corrected for publication bias. Second, the number of completed WCST categories was significantly more affected by PD-related changes than WCST measures of perseveration. Third, WCST deficits were most pronounced in PD patients that were tested after withdrawal from dopaminergic medication and in those samples that were characterized by high disease duration and severe motor impairment. Among these moderators, the degree of motor impairment (as measured by the UPDRS) seems to be the most important predictor of WCST performance deficits in patients with PD.

### 4.1 The size and robustness of WCST deficits in patients with PD

Our observation of significant WCST performance deficits in patients with PD will not be surprising to readers who are familiar with the literature on cognitive impairment in PD. WCST deficits are routinely cited as part of a PD-related pattern of executive dysfunctions (Brown &

Marsden, 1990; Dirnberger & Jahanshahi, 2013; Kehagia, Barker, & Robbins, 2010) and we are not aware of any contemporary doubts about the impairment of WCST performance in PD. However, our meta-analysis revealed novel insights into the size and robustness of these deficits. Most notably, group differences between patients with PD and HC on measures of global WCST performance (i.e., the number of completed categories, the number of total errors) were associated with large effect sizes ($d$ = 0.74 – $d$ = 0.78), which are uncommon in the meta-analytic WCST literature. Substantial WCST deficits have been observed in various neurological and psychiatric disorders, including amyotrophic lateral sclerosis (ALS; Beeldman et al., 2016; Lange, Vogts et al., 2016), primary dystonia (Lange, Seer, Salchow et al., 2016), Gilles de la Tourette syndrome (Lange, Seer, Müller-Vahl, & Kopp, 2017), eating disorders (Roberts, Tchanturia, Stahl, Southgate, & Treasure, 2007), attention deficit hyperactivity disorder (Romine et al., 2004), depression (Snyder, 2013), and obsessive-compulsive disorder (Shin, Lee, Kim, & Kwon, 2014). Across these conditions, disease-related WCST performance deficits are remarkably similar and not larger than medium in size (typically around $d$ = 0.5; Lange, Seer, & Kopp, 2017). For example, a recent meta-analysis on WCST deficits in primary dystonia (Lange, Seer, Salchow et al., 2016) reported an average effect size for the difference between patients and HC in the number of completed categories of $d$ = .41, 95% CI [0.18, 0.64]. Note that the confidence interval around this effect size does not overlap with the corresponding interval from our present analysis of WCST deficits in PD, $d$ = 0.74, 95% CI [0.67, 0.82]. WCST performance deficits on the category measure thus seem to be substantially larger in PD than in primary dystonia. This finding suggests that WCST deficits in PD cannot entirely be attributed to disease-unspecific factors (e.g., symptom-related distraction; Jahanshahi, Rowe, & Fuller, 2003) that are common to all of the conditions listed above. We will return to this possibility when discussing the moderating effect of motor impairment on WCST deficits in PD.

In comparison to an earlier meta-analysis on WCST impairment in PD (Kudlicka et al., 2011), our meta-analysis arrived at more precise effect-size estimates. For example, the 95% confidence interval reported by Kudlicka and colleagues for the PD-related decrease in the number of completed WCST categories ranged from $d = 0.39$ to $d = 0.97$ and was thus almost four times wider than the interval determined in our analysis. Moreover, the large number of studies included in our meta-analysis allowed for a powerful test of the possibility that reported WCST deficits might be inflated by publication bias. Although we found evidence for subtle small-study effects (i.e., statistical relationships between study precision and effect size), corrections for these effects did not substantially alter our results. Deficits with regard to the number of completed WCST categories, for example, remained larger than $d = 0.6$ even when adjusted for the (most likely unrealistic) assumption that reports in the field have been produced under severe one-sided publication bias. The limited influence of publication bias on our meta-analysis may reflect a fortunate decoupling of WCST results and publication success across the included studies. Many of the studies included in our analysis did not exclusively focus on the WCST difference between patients with PD and HC. Authors of these studies administered the WCST as a part of larger batteries of standardized neuropsychological tests or as a background measure when mainly focusing on PD-related alterations in other domains. As a result, the publication of these reports is rather unlikely to depend on statistically significant WCST performance deficits between patients and controls. Our meta-analysis thus also illustrates how the neuropsychological research culture of routinely reporting data from standardized tests can lead to comparatively unbiased literatures and effect-size estimates.

To fully realize this potential, studies involving neuropsychological methodology would benefit from a higher degree of standardization in the reporting of test results. Many studies in the field provide only the names of the administered tests (sometimes without mentioning the test

version and without citation) and do not specify the reported outcome variables (Miller, Schoenberg, & Bilder, 2014). Similarly, it has been noted that for neuropsychological tests involving multiple outcome measures many studies only report an arbitrary selection of outcomes (Loring & Bowden, 2014). We observed both these phenomena when extracting data from studies on WCST performance in patients with PD. To further increase the comparability of neuropsychological studies and the precision of meta-analyses in the field, we would like to encourage the implementation of reporting standards for the presentation of neuropsychological test results. Every study administering the WCST should, for example, explicitly mention the WCST version that was used and report means and standard deviations for all of the outcome measures that can be obtained from this test version. If a study involves a more narrow focus on a particular facet of WCST performance, this focus needs to be justified a priori and an unbiased report of data for all available variables should be given in the supplementary materials.

**4.2 Different facets of WCST performance in patients with PD**

The factors that account for the small but significant difference in the size of PD-related deficits on our two primary WCST measures (i.e., categories and perseverations) cannot be identified with certainty. The difference might result from a statistical artefact (e.g., the category measure might be subject to a ceiling effect that amplifies the group difference) or reflect that the category measure is more sensitive to the type of WCST impairment characteristic for PD. Results from our moderator analysis support the latter possibility as they illustrate that PD-related deficits on the category measure, but not on the perseveration measure, vary as a function of the duration and severity of PD. Importantly, the observed dissociation of WCST performance measures suggests that WCST impairment in PD might not be primarily due to patients' difficulties in the domain of cognitive flexibility. As a complex executive functioning task, the WCST does not exclusively require cognitive flexibility, but also a diverse set of additional

cognitive processes (Buchsbaum et al., 2005; Dehaene & Changeux, 1991; Lange, Seer, & Kopp, 2017; Ridderinkhof, Span, & van der Molen, 2002). Global measures of WCST performance (such as the number of completed categories) reflect the interaction of these processes, while more specific measures (such as the number of perseverations) have the potential to be more process-pure indicators of specific cognitive abilities (e.g., cognitive flexibility). If PD-related WCST impairment were mainly inflexible in nature, we would have expected large deficits on the perseveration measure, which would be diluted (and hence, smaller) in the more global category measure. The fact that we observed the opposite pattern suggests that WCST performance deficits in PD might result from a change in a cognitive process that is more relevant to the number of completed categories than to the number of committed perseverations. Given the available data, this conclusion remains speculative and alternative explanations cannot be excluded. For example, it is also possible that the differential sensitivity of WCST measures to PD-related changes reflects differences in reliability between the measures (Bowden et al., 1998).

A more detailed analysis of WCST performance might allow identifying the cognitive processes that give rise to the decreased number of completed WCST categories in patients with PD. Given the results of our meta-analysis, it might be particularly promising to focus on improving the decomposition of non-perseverative WCST errors (Barceló, 1999; Barceló & Knight, 2002; Barceló, Muñoz-Céspedes, Pozo, & Rubia, 2000; Lange, Kröger, et al. 2016; Nyhus & Barceló, 2009). The non-perseverative error score is an aggregate of all WCST errors that are not perseverative errors. Among others, it confounds failures to maintain set, efficient errors, and integration errors (Lange, Kröger, et al. 2016). In comparison to other WCST measures, PD-related WCST deficits in the number of failures to maintain set seem to be rather small ($d$ = 0.29). Efficient errors occur when participants switch rules after negative feedback, but do not directly identify the newly correct rule. They are necessary to respond flexibly to WCST

task demands and, as a corollary, negatively correlated with the tendency to commit perseverative errors (Godinez, Friedman, Rhee, Miyake, & Hewitt, 2012). Hence, the number of efficient errors can be expected to be smaller rather than larger in patients with PD as compared to HC (i.e., the effect size as scored in our meta-analysis should be $d < 0$). This implies that PD-related deficits with regard to another non-perseverative error type need to be larger than $d = 0.58$ in order for the PD-related deficit in the overall non-perseverative error score to reach the observed effect size of $d = 0.58$. One possible candidate for a type of non-perseverative error that could be disproportionally affected by PD is the so-called integration error (Lange, Kröger, et al. 2016). An integration error is scored when, after an inevitable efficient error, participants fail to integrate the available information to infer the correct new WCST rule. Integration errors are thought to reflect deficient rule-inference processes and have been identified as the primary facet of impairment on a computerized WCST version in older adults and patients with primary dystonia (Lange, Seer, & Kopp, 2017). Separate scoring of integration errors in future studies can reveal to which extent the PD-related increase in non-perseverative errors (and, hence, the decrease in the number of completed WCST categories) is driven by impaired rule inference in patients with PD.

### 4.3 Moderators of WCST performance deficits in patients with PD

Our moderator analyses helped to explain a considerable amount of variability in the size of PD-related WCST deficits across studies. WCST performance deficits were significantly enhanced by the withdrawal of dopaminergic medication and as a function of disease duration and symptom severity. These findings are consistent with a link between progressing striatal dopamine depletion and executive dysfunctions in PD (Cools, Barker, Sahakian, & Robbins, 2003; Leh, Petrides, & Strafella, 2010; MacDonald & Monchi, 2011; Robbins & Cools, 2014). The degeneration of dopaminergic neurons in the substantia nigra pars compacta and the

associated lack of dopamine in the dorsal striatum progress with disease duration (Kordower et al., 2013) and executive functions that involve the dorsal striatum can be expected to follow this trend. Striatal dopamine levels can partially be restored by dopamine replacement therapy, which may relate to a corresponding improvement of executive functioning in PD. Note, however, that the link between disease duration, dopaminergic medication, and WCST performance deficits demonstrated in our meta-analysis does not necessarily imply that striatal dopamine plays a role in the cognitive processes underlying WCST performance. Disease duration and withdrawal of dopaminergic medication are also associated with exacerbated motor symptoms. The severity of motor impairment emerged as an additional predictor in our moderator analysis and, in contrast to disease duration and medication status, it was the only moderator that explained unique variance in the size of WCST performance deficits. WCST impairment in patients with PD thus seems to primarily vary as a function of motor impairment. Motor symptoms have been proposed to constitute a distraction during neuropsychological testing, which can affect patients' cognitive performance (Jahanshahi et al., 2003, 2014). Rather than resulting entirely from underlying neuropathological changes to dopaminergic systems, WCST performance deficits in PD may at least partly be caused by symptom-related distraction. Similarly, the effects of disease duration and medication status on WCST performance might be mediated through their influence on patients' motor symptoms. Future studies are needed to manipulate dopaminergic status while carefully controlling the effects of symptom-related distraction to dissociate primary and secondary contributions to WCST performance deficits in PD.

The presence of substantial WCST performance deficits in never-medicated *de novo* patients with PD and in patients in the lowest UPDRS quartile further supports the generality of this neuropsychological symptom in PD. Likewise, WCST performance was found to be impaired in those studies that explicitly excluded patients with dementia or depression. Hence,

impaired WCST performance in patients with PD seems to be a highly robust phenomenon that can be observed across a large range of patient characteristics.

### 4.4 Future directions

Although our moderator analysis offered some tentative insights into the factors that contribute to WCST performance deficits in PD, it does not allow drawing definitive inferences with regard to mechanisms underlying this neuropsychological symptom. More studies relating WCST performance in PD to neurophysiological data (Cropley et al., 2008; Gawrys et al., 2014; Jubault, Monetta, Strafella, Lafontaine, & Monchi, 2009; Lange, Seer, Loens, et al., 2016; Monchi et al., 2004; Monchi, Petrides, Mejia-Constain, & Strafella, 2007; Nagano-Saito et al., 2014) are required to characterize the neural substrates of WCST impairment in PD. In addition, studies evaluating the impact of deep-brain stimulation (e.g., Jahanshahi et al., 2000; Martínez-Martínez, Aguilar, & Acevedo-Triana, 2017) or dopaminergic medication (e.g., Gotham, Brown, & Marsden, 1988; Pascual-Sedano et al., 2008) might offer more direct evidence with regard to causal relationships between neural changes and WCST deficits in patients with PD. Finally, it would be desirable if more studies compared WCST performance in PD not only to HC but also to a clinical control group (e.g., Cordato, Halliday, Caine, & Morris, 2006; Dujardin, Defebvre, Krystkowiak, Degreef, & Destee, 2003; Puertas-Martín, et al., 2016). Demonstrating PD-related WCST impairment in contrast to a group of patients with comparable motor symptoms would support a link between the pathophysiology of PD and cognitive inflexibility that cannot be attributed to disease-unspecific factors (e.g., symptom-related distraction; cf. Lange, Seer, Dengler, Dressler, & Kopp, 2016).

## 5 Conclusion

PD is associated with robust performance deficits on the WCST. These deficits can also be observed in non-demented, non-depressed, and never-medicated patients with PD, and they

are linked to the severity of patients' motor symptoms. Given the large number of studies providing evidence in support of this change, altered WCST performance can be considered a well-established neuropsychological symptom in patients with PD.

**Acknowledgements**

Florian Lange and Carolin Brückner were supported by the German National Academic Foundation. In addition, Florian Lange received funding from KU Leuven Internal Funds as well as from the FWO and European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665501. Carolin Brückner received an Erasmus Placement scholarship and Bruno Kopp was supported by a grant from the Petermax-Müller-Stiftung, Hannover, Germany.

## References

References marked with an asterisk (*) indicate articles included in the meta-analysis.


*Abo-El-Naga, Y. (2006). Cognitive profile of patients with newly diagnosed Parkinson's. *Current Psychiatry*, *13*(3), 460-473.

*Agosta, F., Gatti, R., Sarasso, E., Volonté, M. A., Canu, E., Meani, A., ... & Comi, G. (2017). Brain plasticity in Parkinson's disease with freezing of gait induced by action observation training. *Journal of Neurology*, *264*(1), 88-101.

*Alevriadou, A., Katsarou, Z., Bostantjopoulou, S., Kiosseoclou, G., & Mentenopoulos, G. (1999). Wisconsin Card Sorting Test variables in relation to motor symptoms in Parkinson's disease. *Perceptual and Motor Skills*, *89*(3), 824-830.

*Alonso Recio, L., Martín, P., Carvajal, F., Ruiz, M., & Serrano, J. M. (2013). A holistic analysis of relationships between executive function and memory in Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *35*(2), 147-159.

*Asahina, M., Suhara, T., Shinotoh, H., Inoue, O., Suzuki, K., & Hattori, T. (1998). Brain muscarinic receptors in progressive supranuclear palsy and Parkinson's disease: a positron emission tomographic study. *Journal of Neurology, Neurosurgery & Psychiatry*, *65*(2), 155-163.

*Assogna, F., Pontieri, F. E., Cravello, L., Peppe, A., Pierantozzi, M., Stefani, A., ... & Spalletta, G. (2010). Intensity-dependent facial emotion recognition and cognitive functions in Parkinson's disease. *Journal of the International Neuropsychological Society*, *16*(5), 867-876.

*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*(12), 1115-1120.

*Baran, B., Tekcan, A. I., Gürvit, H., & Boduroglu, A. (2009). Episodic memory and metamemory in Parkinson's disease patients. *Neuropsychology*, *23*(6), 736 -745.

Barceló, F. (1999). Electrophysiological evidence of two different types of error in the Wisconsin Card Sorting Test. *NeuroReport*, *10*, 1299-1303.

Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*, *40*, 349-356.

Barceló, F., Muñoz-Céspedes, J. M., Pozo, M. A., & Rubia, F. J. (2000). Attentional set shifting modulates the target P3b response in the Wisconsin card sorting test. *Neuropsychologia*, *38*, 1342-1355.

*Beatty, W. W., & Monson, N. (1990). Problem solving in Parkinson's disease: Comparison of performance on the Wisconsin and California Card Sorting Tests. *Topics in Geriatrics*, *3*(3), 163-171.

*Beatty, W. W., Staton, R. D., Weir, W. S., Monson, N., & Whitaker, H. A. (1989). Cognitive disturbances in Parkinson's disease. *Topics in Geriatrics*, *2*(1), 22-33.

Beeldman, E., Raaphorst, J., Twennaar, M. K., de Visser, M., Schmand, B. A., & de Haan, R. J. (2016). The cognitive profile of ALS: a systematic review and meta-analysis update. *Journal of Neurology, Neurosurgery & Psychiatry*, *87*, 611-619.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, *39*, 15–22.

*Blonder, L. X., Gur, R. E., Gur, R. C., Saykin, A. J., & Hurtig, H. I. (1989). Neuropsychological functioning in hemiparkinsonism. *Brain and Cognition*, *9*(2), 244-257.

*Bokura, H., Yamaguchi, S., & Kobayashi, S. (2005). Event-related potentials for response inhibition in Parkinson's disease. *Neuropsychologia*, *43*(6), 967-975.

*Borghammer, P., Østergaard, K., Cumming, P., Gjedde, A., Rodell, A., Hall, N., & Chakravarty, M. M. (2010). A deformation-based morphometry study of patients with early-stage Parkinson's disease. *European Journal of Neurology*, *17*(2), 314-320.

Bowden, S. C., Fowler, K. S., Bell, R. C., Whelan, G., Clifford, C. C., Ritter, A. J., & Long, C. M. (1998). The reliability and internal validity of the Wisconsin Card Sorting Test. *Neuropsychological Rehabilitation*, *8*(3), 243-254.

*Brand, M., Labudda, K., Kalbe, E., Hilker, R., Emmans, D., Fuchs, G., ... & Markowitsch, H. J. (2004). Decision-making impairments in patients with Parkinson's disease. *Behavioural Neurology*, *15*(3-4), 77-85.

*Breitenstein, C., Van Lancker, D., Daum, I., & Waters, C. H. (2001). Impaired perception of vocal emotions in Parkinson's disease: influence of speech time processing and executive functioning. *Brain and Cognition*, *45*(2), 277-314.

*Broeders, M., Velseboer, D. C., de Bie, R., Speelman, J. D., Muslimovic, D., Post, B., ... & Schmand, B. (2013). Cognitive change in newly-diagnosed patients with Parkinson's disease: a 5-year follow-up study. *Journal of the International Neuropsychological Society*, *19*(6), 695-708.

*Broussolle, E., Dentresangle, C., Landais, P., Garcia-Larrea, L., Pollak, P., Croisile, B., ... & Comar, D. (1999). The relation of putamen and caudate nucleus 18 F-Dopa uptake to motor and cognitive performances in Parkinson's disease. *Journal of the Neurological Sciences*, *166*(2), 141-151.

*Brown, G. G., Brown, S. J., Christenson, G., Williams, R. E., Kindermann, S. S., Loftis, C., ... & Gorell, J. M. (2002). Effects of task structure on category priming in patients with Parkinson's disease and in healthy individuals. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 356-369.

*Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain*, *111*(2), 323-345.

*Brown, R. G., & Marsden, C. D. (1988b). An investigation of the phenomenon of "set" in Parkinson's disease. *Movement Disorders*, *3*(2), 152-161.

Brown, R. G., & Marsden, C. D. (1990). Cognitive function in Parkinson's disease: from description to theory. *Trends in Neurosciences*, *13*(1), 21-29.

Buchsbaum, B. R., Greer, S., Chang, W. L., & Berman, K. F. (2005). Meta-analysis of neuroimaging studies of the Wisconsin card sorting task and component processes. *Human Brain Mapping*, *25*(1), 35-45.

*Caffarra, P., Majer, S., Dieci, F., Messa, G., Scaglioni, A., & Salmaso, D. (2012). *Organizzazione della memoria nella malattia di Parkinson* [PDF document]. Retrieved from: http://www.ds3web.it/papers/ItalianNetworkAD_1999.pdf

*Caltagirone, C., Carlesimo, A., Nocentini, U., & Vicari, S. (1989). Defective concept formation in parkinsonians is independent from mental deterioration. *Journal of Neurology, Neurosurgery & Psychiatry*, *52*(3), 334-337.

*Campos-Sousa, I. S., Campos-Sousa, R. N., Ataíde Jr, L., Soares, M. M. D. B., & Almeida, K. J. (2010). Executive dysfunction and motor symptoms in Parkinson's disease. *Arquivos de Neuro-psiquiatria*, *68*(2), 246-251.

*Canavan, A. G. M., Passingham, R. E., Marsden, C. D., Quinn, N., Wyke, M., & Polkey, C. E. (1989). The performance on learning tasks of patients in the early stages of Parkinson's disease. *Neuropsychologia*, *27*(2), 141-156.

*Canu, E., Agosta, F., Sarasso, E., Volontè, M. A., Basaia, S., Stojkovic, T., ... & Gatti, R. (2015). Brain structural and functional connectivity in Parkinson's disease with freezing of gait. *Human Brain Mapping*, *36*(12), 5064-5078.

*Cerasa, A., Salsone, M., Nigro, S., Chiriaco, C., Donzuso, G., Bosco, D., ... & Quattrone, A. (2014). Cortical volume and folding abnormalities in Parkinson's disease patients with pathological gambling. *Parkinsonism & Related Disorders*, *20*(11), 1209-1214.

*Chang, Y. P., Yang, Y. H., Lai, C. L., & Liou, L. M. (2016). Event-Related Potentials in Parkinson's Disease Patients with Visual Hallucination. *Parkinson's Disease, 2016:* 1863508.

*Chau, P. M. (2010). *Laterality of Motor Symptom Onset, Disease Progression, and Cognition in Parkinson's Disease* (Unpublished doctoral dissertation). University of Kansas, Kansas.

*Chen, K. H., Okerstrom, K. L., Kingyon, J. R., Anderson, S. W., Cavanagh, J. F., & Narayanan, N. S. (2016). Startle habituation and midfrontal theta activity in Parkinson disease. *Journal of Cognitive Neuroscience, 28(12),* 1923-1932.

*Chen, K. J., Lin, R. T., Liu, C. K., Tai, C. T., & Lai, C. L. (2006). Relationship between event-related potentials and frontal-subcortical dysfunction in Parkinson's disease. *Parkinsonism & Related Disorders*, *12*(7), 453-458.

Christopher, L., & Strafella, A. P. (2013). Neuroimaging of brain changes associated with cognitive impairment in Parkinson's disease. *Journal of Neuropsychology, 7*(2), 225–240.

*Clark, L. (2014). *Relation of Personality Traits to Cognitive Impairments and Disease Severity in Parkinson's Disease* (Unpublished thesis). Bridgewater State University, Bridgewater.

*Cohn, M., Giannoylis, I., De Belder, M., Saint-Cyr, J. A., & McAndrews, M. P. (2016). Associative reinstatement memory measures hippocampal function in Parkinson's Disease. *Neuropsychologia*, *90*, 25-32.

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2003). L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinson's disease. *Neuropsychologia*, *41*(11), 1431-1441.

762 *Cooper, J. A., Sagar, H. J., Jordan, N., Harvey, N. S., & Sullivan, E. V. (1991). Cognitive
763 impairment in early, untreated Parkinson's disease and its relationship to motor
764 disability. *Brain*, *114*(5), 2095-2122.
765 *Cordato, N. J., Halliday, G. M., Caine, D., & Morris, J. G. (2006). Comparison of motor,
766 cognitive, and behavioral features in progressive supranuclear palsy and Parkinson's
767 disease. *Movement Disorders*, *21*(5), 632-638.
768 *Costa, A., Peppe, A., Carlesimo, G. A., Salamone, G., & Caltagirone, C. (2007).
769 Neuropsychological correlates of alexithymia in Parkinson's disease. *Journal of the*
770 *International Neuropsychological Society*, *13*(6), 980-992.
771 *Costa, A., Zabberoni, S., Peppe, A., Serafini, F., Scalici, F., Caltagirone, C., & Carlesimo, G. A.
772 (2015). Time-based prospective memory functioning in mild cognitive impairment associated
773 with Parkinson's disease: relationship with autonomous management of daily living
774 commitments. *Frontiers in Human Neuroscience*, *9: 333.*
775 *Crescentini, C., Marin, D., Del Missier, F., Biasutti, E., & Shallice, T. (2011). Interference from
776 retrieval cues in Parkinson's disease. *Neuropsychology*, *25*(6), 720-733.
777 *Crescentini, C., Mondolo, F., Biasutti, E., & Shallice, T. (2012). Preserved and impaired task-
778 switching abilities in non-demented patients with Parkinson's disease. *Journal of*
779 *Neuropsychology*, *6*(1), 94-118.
780 *Cropley, V. L., Fujita, M., Bara-Jimenez, W., Brown, A. K., Zhang, X. Y., Sangare, J., ... &
781 Innis, R. B. (2008). Pre-and post-synaptic dopamine imaging and its relation with
782 frontostriatal cognitive function in Parkinson disease: PET studies with [11 C] NNC 112 and
783 [18 F] FDOPA. *Psychiatry Research: Neuroimaging*, *163*(2), 171-182.

*Dalrymple-Alford, J. C., Jamieson, C. F., & Donaldson, I. M. (1995). Effects of selegiline (deprenyl) on cognition in early Parkinson's disease. *Clinical Neuropharmacology*, *18*(4), 348-359.

*Dalrymple-Alford, J. C., Kalders, A. S., Jones, R. D., & Watson, R. W. (1994). A central executive deficit in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(3), 360-367.

*Davidson, P. S., Anaki, D., Saint-Cyr, J. A., Chow, T. W., & Moscovitch, M. (2006). Exploring the recognition memory deficit in Parkinson's disease: estimates of recollection versus familiarity. *Brain*, *129*(7), 1768-1779.

*Davidson, P. S., Cook, S. P., McGhan, L., Bouchard, T., & Camicioli, R. (2013). Source memory in normal aging and Parkinson's disease. *Journal of Neuropsychology*, *7*(2), 179-192.

Dehaene, S., & Changeux, J. P. (1991). The Wisconsin Card Sorting Test: theoretical analysis and modelling in a neuronal network. *Cerebral Cortex*, *1*, 62-79.

Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*, *17*, 255-264.

*Díaz-Santos, M., Cao, B., Yazdanbakhsh, A., Norton, D. J., Neargarder, S., & Cronin-Golomb, A. (2015). Perceptual, cognitive, and personality rigidity in Parkinson's disease. *Neuropsychologia*, *69*, 183-193.

Dirnberger, G., & Jahanshahi, M. (2013). Executive dysfunction in Parkinson's disease: a review. *Journal of Neuropsychology*, *7*(2), 193-224.

*Doyon, J., Bourgeois, C., & Bédard, P. (1996). Déficits visuo-spatiaux associés à la Maladie de Parkinson. *International Journal of Psychology*, *31*(5), 161-175.

*Drag, L. L., Bieliauskas, L. A., Kaszniak, A. W., Bohnen, N. I., & Glisky, E. L. (2009). Source memory and frontal functioning in Parkinson's disease. *Journal of the International Neuropsychological Society*, *15*(3), 399-406.

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

*Dubois, B., Pillon, B., Legault, F., Agid, Y., & Lhermitte, F. (1988). Slowing of cognitive processing in progressive supranuclear palsy: a comparison with Parkinson's disease. *Archives of Neurology*, *45*(11), 1194-1199.

*Dubois, B., Pillon, B., Sternic, N., Lhermitte, F., & Agid, Y. (1990). Age-induced cognitive disturbances in Parkinson's disease. *Neurology*, *40*(1), 38-38.

*Dujardin, K., Defebvre, L., Grunberg, C., Becquet, E., & Destée, A. (2001). Memory and executive function in sporadic and familial Parkinson's disease. *Brain*, *124*(2), 389-398.

*Dujardin, K., Defebvre, L., Krystkowiak, P., Degreef, J. F., & Destee, A. (2003). Executive function differences in multiple system atrophy and Parkinson's disease. *Parkinsonism & Related Disorders*, *9*(4), 205-211.

*Ebmeier, K. P., Potter, D. D., Cochrane, R. H., Crawford, J. R., Stewart, L., Calder, S. A., ... & Salzen, E. A. (1992). Event related potentials, reaction time, and cognitive performance in idiopathic Parkinson's disease. *Biological Psychology*, *33*(1), 73-89.

*Ekman, U., Eriksson, J., Forsgren, L., Mo, S. J., Riklund, K., & Nyberg, L. (2012). Functional brain activity and presynaptic dopamine uptake in patients with Parkinson's disease and mild cognitive impairment: a cross-sectional study. *The Lancet Neurology*, *11*(8), 679-687.

*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.

Elliott, R. (2003). Executive functions and their disorders: Imaging in clinical neuroscience. *British Medical Bulletin*, *65*(1), 49-59.

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

*Euteneuer, F., Schaefer, F., Stuermer, R., Boucsein, W., Timmermann, L., Barbe, M. T., ... & Kalbe, E. (2009). Dissociation of decision-making under ambiguity and decision-making under risk in patients with Parkinson's disease: a neuropsychological and psychophysiological study. *Neuropsychologia*, *47*(13), 2882-2890.

*Fales, C. L., Vanek, Z. F., & Knowlton, B. J. (2006). Backward inhibition in Parkinson's disease. *Neuropsychologia*, *44*(7), 1041-1049.

*Fama, R., Sullivan, E. V., Shear, P. K., Stein, M., Yesavage, J. A., Tinklenberg, J. R., & Pfefferbaum, A. (2000). Extent, pattern, and correlates of remote memory impairment in Alzheimer's disease and Parkinson's disease. *Neuropsychology*, *14*(2), 265-276.

*Farina, E., Gattellaro, G., Pomati, S., Magni, E., Perretti, A., Cannata, A. P., ... & Mariani, C. (2000). Researching a differential impairment of frontal functions and explicit memory in early Parkinson's disease. *European Journal of Neurology*, *7*(3), 259-267.

Field, A. P. (2013). Discovering statistics using IBM SPSS statistics. London, UK: Sage.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 665–694.

*Filoteo, J. V., Maddox, W. T., Ing, A. D., Zizak, V., & Song, D. D. (2005). The impact of irrelevant dimensional variation on rule-based category learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society*, *11*(5), 503-513.

*Flensborg Damholdt, M., Shevlin, M., Borghammer, P., Larsen, L., & Østergaard, K. (2012). Clinical heterogeneity in Parkinson's disease revisited: a latent profile analysis. *Acta Neurologica Scandinavica*, *125*(5), 311-318.

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.

*Fonoff, F. C., Fonoff, E. T., Barbosa, E. R., Quaranta, T., Machado, R. B., de Andrade, D. C., ... & Fuentes, D. (2015). Correlation between impulsivity and executive function in patients with Parkinson disease experiencing depression and anxiety symptoms. *Journal of Geriatric Psychiatry and Neurology*, *28*(1), 49-56.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186-204.

*Galtier, I., Nieto, A., Lorenzo, J. N., & Barroso, J. (2014). Cognitive impairment in Parkinson's disease: more than a frontostriatal dysfunction. *The Spanish Journal of Psychology*, *17*(68), 1-8.

*Gasparini, M., Bonifati, V., Fabrizio, E., Fabbrini, G., Brusa, L., Lenzi, G. L., & Meco, G. (2001). Frontal lobe dysfunction in essential tremor. *Journal of Neurology*, *248*(5), 399-402.

874 *Gauggel, S., Rieger, M., & Feghoff, T. A. (2004). Inhibition of ongoing responses in patients
875 with Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(4), 539-544.
876 *Gawrys, L., Falkiewicz, M., Pilacinski, A., Riegel, M., Piatkowska-Janko, E., Bogorodzki, P., ...
877 & Koziorowski, D. (2014). The neural correlates of specific executive dysfunctions in
878 Parkinson's disease. *Acta Neurobiol Exp*, *74*, 465-478.
879 *Gawrys, L., Szatkowska, I., Jamrozik, Z., Janik, P., Friedman, A., & Kaczmarek, L. (2008).
880 Nonverbal deficits in explicit and implicit memory of Parkinson's disease patients. *Acta
881 Neurobiologiae Experimentalis*, *68*(1), 58-72.
882 *Gnanalingham, K. K., Byrne, E. J., Thornton, A., Sambrook, M. A., & Bannister, P. (1997).
883 Motor and cognitive function in Lewy body dementia: comparison with Alzheimer9s and
884 Parkinson9s diseases. *Journal of Neurology, Neurosurgery & Psychiatry*, *62*(3), 243-252.
885 Godinez, D. A., Friedman, N. P., Rhee, S. H., Miyake, A., & Hewitt, J. K. (2012). Phenotypic
886 and genetic analyses of the Wisconsin Card Sort. *Behavior Genetics*, *42*(2), 209-220.
887 *Gotham, A. M., Brown, R. G., & Marsden, C. D. (1988). 'Frontal' cognitive function in patients
888 with Parkinson's disease 'on' and 'off' levodopa. *Brain*, *111*(2), 299-321.
889 *Graham, J. M., Paley, M. N., Grünewald, R. A., Hoggard, N., & Griffiths, P. D. (2000). Brain
890 iron deposition in Parkinson's disease imaged using the PRIME magnetic resonance
891 sequence. *Brain*, *123*(12), 2423-2431.
892 Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of
893 shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental
894 Psychology*, *38*(4), 404–411.
895 *Hanby, M. F., Barraclough, M., McKie, S., Hinvest, N., McDonald, K., Elliott, R., & Leroi, I.
896 (2014). Emotional and Cognitive Processing Deficits in People with Parkinson's Disease and
897 Apathy. *Journal of Alzheimer's Disease & Parkinsonism*, *4*(5): 1000156.

*Hawkins, K. A., Jennings, D., Vincent, A. S., Gilliland, K., West, A., & Marek, K. (2012). Traditional neuropsychological correlates and reliability of the automated neuropsychological assessment metrics-4 battery for Parkinson's disease. *Parkinsonism & Related Disorders, 18*(7), 864-870.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 362*(1485), 1601–1613.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test. Professional manual*. Lutz, FL: Psychological Assessment Resources.

Herd, S. A., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014). A neural network model of individual differences in task switching abilities. *Neuropsychologia, 62,* 375-389.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The BMJ, 327*(7414), 557–560.

*Hocherman, S., Moont, R., & Schwartz, M. (2004). Response selection and execution in patients with Parkinson's disease. *Cognitive Brain Research, 19*(1), 40-51.

*Hozumi, A., Hirata, K., Tanaka, H., & Yamazaki, K. (2000). Perseveration for novel stimuli in Parkinson's disease: An evaluation based on event-related potentials topography. *Movement Disorders, 15*(5), 835-842.

*Iijima, M., Osawa, M., Iwata, M., Miyazaki, A., & Tei, H. (2000). Topographic mapping of P300 and frontal cognitive function in Parkinson's disease. *Behavioural Neurology, 12*(3), 143-148.

*Inzelberg, R., Plotnik, M., Flash, T., Schechtman, E., Shahar, I., & Korczyn, A. D. (2001). Mental and motor switching in Parkinson's disease. *Journal of Motor Behavior*, *33*(4), 377-385.

*Ito, J., & Kitagawa, J. (2006). Performance monitoring and error processing during a lexical decision task in patients with Parkinson's disease. *Journal of Geriatric Psychiatry and Neurology*, *19*(1), 46-54.

Jahanshahi, M., Ardouin, C. M. A., Brown, R. G., Rothwell, J. C., Obeso, J., Albanese, A., ... & Limousin-Dowsey, P. (2000). The impact of deep brain stimulation on executive function in Parkinson's disease. *Brain*, *123*(6), 1142-1154.

Jahanshahi, M., Rowe, J., & Fuller, R. (2003). Cognitive executive function in dystonia. *Movement Disorders*, *18*(12), 1470–1481.

*Jahanshahi, M., Rowe, J., Saleem, T., Brown, R. G., Limousin-Dowsey, P., Rothwell, J. C., ... & Quinn, N. P. (2002). Striatal contribution to cognition: working memory and executive function in Parkinson's disease before and after unilateral posteroventral pallidotomy. *Journal of Cognitive Neuroscience*, *14*(2), 298-310.

Jahanshahi, M., Torkamani, M., Beigi, M., Wilkinson, L., Page, D., Madeley, L.,. . . Tisch, S. (2014). Pallidal stimulation for primary generalised dystonia: effect on cognition, mood and quality of life. *Journal of Neurology*, *261*(1), 164–173.

Janvin, C. C., Aarsland, D., & Larsen, J. P. (2005). Cognitive predictors of dementia in Parkinson's disease: a community-based, 4-year longitudinal study. *Journal of Geriatric Psychiatry and Neurology*, *18*(3), 149-154.

Jubault, T., Monetta, L., Strafella, A. P., Lafontaine, A.-L., Monchi, O., & Gribble, P. L. (2009). L-Dopa medication in Parkinson's disease restores activity in the motor cortico-striatal loop but does not modify the cognitive network. *PLoS ONE*, *4*(7), e6154.

945 *Katai, S., Maruyama, T., Hashimoto, T., & Ikeda, S. (2003). Event based and time based
946 prospective memory in Parkinson's disease. *Journal of Neurology, Neurosurgery &*
947 *Psychiatry*, *74*(6), 704-709.
948 *Katsarou, Z., Bostantjopoulou, S., Kimiskidis, V., Rossopoulos, E., & Kazis, A. (2004).
949 Auditory event-related potentials in Parkinson's disease in relation to cognitive
950 ability. *Perceptual and Motor Skills*, *98*, 1441-1448.
951 *Kaufman, D. A., Bowers, D., Okun, M. S., Van Patten, R., & Perlstein, W. M. (2016). Apathy,
952 novelty Processing, and the P3 Potential in Parkinson's Disease. *Frontiers in Neurology*, *7*:
953 95.
954 Kehagia, A. A., Barker, R. A., & Robbins, T. W. (2010). Neuropsychological and clinical
955 heterogeneity of cognitive impairment and dementia in patients with Parkinson's disease. *The*
956 *Lancet Neurology*, *9*(12), 1200-1213.
957 Kehagia, A. A., Barker, R. A., & Robbins, T. W. (2013). Cognitive impairment in Parkinson's
958 disease: the dual syndrome hypothesis. *Neurodegenerative Diseases*, *11*(2), 79-92.
959 Kordower, J. H., Olanow, C. W., Dodiya, H. B., Chu, Y., Beach, T. G., Adler, C. H., … Bartus,
960 R. T. (2013). Disease duration and the integrity of the nigrostriatal system in Parkinson's
961 disease. *Brain*, *136*(8), 2419–2431.
962 *Krishna, R., Ali, M., & Moustafa, A. A. (2014). Effects of combined MAO-B inhibitors and
963 levodopa vs. monotherapy in Parkinson's disease. *Frontiers in Aging Neuroscience*, *6*: 180.
964 Kudlicka, A., Clare, L., & Hindle, J. V. (2011). Executive functions in Parkinson's disease:
965 systematic review and meta-analysis. *Movement Disorders*, *26*(13), 2305–2315.
966 Kudlicka, A., Clare, L., & Hindle, J. V. (2014). Quality of life, health status and caregiver burden
967 in Parkinson's disease: relationship to executive functioning. *International Journal of*
968 *Geriatric Psychiatry*, *29*(1), 68-76.

*Labudda, K., Brand, M., Mertens, M., Ollech, I., Markowitsch, H. J., & Woermann, F. G. (2010). Decision making under risk condition in patients with Parkinson's disease: a behavioural and fMRI study. *Behavioural Neurology*, *23*(3), 131-143.

Lange, F., Kröger, B., Steinke, A., Seer, C., Dengler, R., & Kopp, B. (2016). Decomposing card-sorting performance: Effects of working memory load and age-related changes. *Neuropsychology*, *30*(5), 579-590.

Lange, F., Seer, C., & Kopp, B. (2017). Cognitive flexibility in neurological disorders: Cognitive components and event-related potentials. *Neuroscience & Biobehavioral Reviews*, *83*, 496-507.

Lange, F., Seer, C., Dengler, R., Dressler, D., & Kopp, B. (2016). Cognitive flexibility in primary dystonia. *Journal of the International Neuropsychological Society*, *22*(6), 662-670.

*Lange, F., Seer, C., Loens, S., Wegner, F., Schrader, C., Dressler, D., ... & Kopp, B. (2016). Neural mechanisms underlying cognitive inflexibility in Parkinson's disease. *Neuropsychologia*, *93*, 142-150.

Lange, F., Seer, C., Müller-Vahl, K., & Kopp, B. (2017). Cognitive flexibility and its electrophysiological correlates in Gilles de la Tourette syndrome. *Developmental Cognitive Neuroscience*, *27*, 78-90.

Lange, F., Seer, C., Salchow, C., Dengler, R., Dressler, D., & Kopp, B. (2016). Meta-analytical and electrophysiological evidence for executive dysfunction in primary dystonia. *Cortex*, *82*, 133-146.

Lange, F., Vogts, M.-B., Seer, C., Fürkötter, S., Abdulla, S., Dengler, R., Kopp, B., & Petri, S. (2016). Impaired set-shifting in amyotrophic lateral sclerosis: An event-related potential study of executive function. *Neuropsychology*, *30*(1), 120-134.

Lees, A. J. (1989). The neurobehavioural abnormalities in Parkinson's disease and their relationship to psychomotor retardation and obsessional compulsive disorders. *Behavioural Neurology*, *2*(1), 1-11.

*Lees, A. J., & Smith, E. (1983). Cognitive deficits in the early stages of Parkinson's disease. *Brain*, *106*(2), 257-270.

Leh, S. E., Petrides, M., & Strafella, A. P. (2010). The neural circuitry of executive functions in healthy subjects and Parkinson's disease. *Neuropsychopharmacology*, *35*(1), 70–85.

Lenhard, W. & Lenhard, A. (2016). Berechnung von Effektstärken. verfügbar unter: https://www.psychometrica.de/effektstaerke.html. Dettelbach: Psychometrica.

*Leroi, I., Barraclough, M., McKie, S., Hinvest, N., Evans, J., Elliott, R., & McDonald, K. (2013). Dopaminergic influences on executive function and impulsive behaviour in impulse control disorders in Parkinson's disease. *Journal of Neuropsychology*, *7*(2), 306-325.

*Leroi, I., Pantula, H., McDonald, K., & Harbishettar, V. (2012). Neuropsychiatric symptoms in Parkinson's disease with mild cognitive impairment and dementia. *Parkinson's Disease, 2012:* 308097.

*Levin, B. E., Llabre, M. M., & Weiner, W. J. (1989). Cognitive impairments associated with early Parkinson's disease. *Neurology*, *39*(4), 557-557.

*Liozidou, A., Potagas, C., Papageorgiou, S. G., & Zalonis, I. (2012). The role of working memory and information processing speed on Wisconsin card sorting test performance in Parkinson disease without dementia. *Journal of Geriatric Psychiatry and Neurology*, *25*(4), 215-221.

*Lohmann, E., Thobois, S., Lesage, S., Broussolle, E., du Montcel, S. T., Ribeiro, M. J., ... & Pollak, P. (2009). A multidisciplinary study of patients with early-onset PD with and without parkin mutations. *Neurology*, *72*(2), 110-116.

Loring, D. W., & Bowden, S. C. (2014). The STROBE statement and neuropsychology: lighting the way toward evidence-based practice. *The Clinical Neuropsychologist*, *28*(4), 556-574.

MacDonald, P. A., & Monchi, O. (2011). Differential effects of dopaminergic therapies on dorsal and ventral striatum in Parkinson's disease: implications for cognitive function. *Parkinson's Disease*, *2011*: 572743.

Mahieux, F., Fénelon, G., Flahault, A., Manifacier, M. J., Michelet, D., & Boller, F. (1998). Neuropsychological prediction of dementia in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *64*(2), 178-183.

*Marklund, P., Larsson, A., Elgh, E., Linder, J., Riklund, K. Å., Forsgren, L., & Nyberg, L. (2008). Temporal dynamics of basal ganglia under-recruitment in Parkinson's disease: transient caudate abnormalities during updating of working memory. *Brain*, *132*(2), 336-346.

Martínez-Martínez, A. M., Aguilar, O. M., & Acevedo-Triana, C. A. (2017). Meta-Analysis of the Relationship between Deep Brain Stimulation in Patients with Parkinson's Disease and Performance in Evaluation Tests for Executive Brain Functions. *Parkinson's Disease*, *2017*: 9641392.

*McDowd, J., Hoffman, L., Rozek, E., Lyons, K. E., Pahwa, R., Burns, J., & Kemper, S. (2011). Understanding verbal fluency in healthy aging, Alzheimer's disease, and Parkinson's disease. *Neuropsychology*, *25*(2), 210-225.

*Mignard, P., Cave, C., Lagrue, B., Meynadier, Y., & Viallet, F. (2001). Etude instrumentale des pauses silencieuses chez les parkinsoniens en production orale spontanée et en lecture. *Revue de Neuropsychologie*, *11*(1), 39-63.

Miller, I. N., Neargarder, S., Risi, M. M., & Cronin–Golomb, A. (2013). Frontal and posterior subtypes of neuropsychological deficit in Parkinson's disease. *Behavioral Neuroscience*, *127*(2), 175-183.

Miller, J. B., Schoenberg, M. R., & Bilder, R. M. (2014). Consolidated Standards of Reporting Trials (CONSORT): considerations for neuropsychological research. *The Clinical Neuropsychologist*, *28*(4), 575-599.

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

*Mioni, G., Meligrana, L., Grondin, S., Perini, F., Bartolomei, L., & Stablum, F. (2016). Effects of emotional facial expression on time perception in patients with Parkinson's disease. *Journal of the International Neuropsychological Society*, *22*(9), 890-899.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.

*Mohr, E., Juncos, J., Cox, C., Litvan, I., Fedio, P., & Chase, T. N. (1990). Selective deficits in cognition and memory in high-functioning parkinsonian patients. *Journal of Neurology, Neurosurgery & Psychiatry*, *53*(7), 603-606.

*Mollion, H., Ventre-Dominey, J., Dominey, P. F., & Broussolle, E. (2003). Dissociable effects of dopaminergic therapy on spatial versus non-spatial working memory in Parkinson's disease. *Neuropsychologia*, *41*(11), 1442-1451.

Monchi, O., Hanganu, A., & Bellec, P. (2016). Markers of cognitive decline in PD: the case for heterogeneity. *Parkinsonism & Related Disorders*, *24*, 8–14.

Monchi, O., Petrides, M., Doyon, J., Postuma, R. B., Worsley, K., & Dagher, A. (2004). Neural bases of set-shifting deficits in Parkinson's disease. *Journal of Neuroscience*, *24*(3), 702-710.

Monchi, O., Petrides, M., Mejia-Constain, B., & Strafella, A. P. (2007). Cortical activity in Parkinson's disease during executive processing depends on striatal involvement. *Brain*, *130*(1), 233–244.

*Moro dos Santos, C., Bisol Balardin, J., Quarti Irigaray, T., Schröder, N., Rieder, C. R., & Bromberg, E. (2010). Incidental encoding strategies did not improve contextual memory in Parkinson's disease patients. *Neurorehabilitation and Neural Repair*, *24*(5), 450-456.

*Müller, U., Wächter, T., Barthel, H., Reuter, M., & von Cramon, D. Y. (2000). Striatal [123 I] β-CIT SPECT and prefrontal cognitive functions in Parkinson's disease. *Journal of Neural Transmission*, *107*(3), 303-319.

*Muñiz Casado, J. A., & Osuna Benavides, A. (2007). Estudio de las funciones ejecutivas en pacientes con enfermedad de Parkinson en fases iniciales. *Mapfre Medicina*, *18*(1), 46-53.

*Münte, T. F., Joppich, G., Däuper, J., Schrader, C., Dengler, R., & Heldmann, M. (2015). Random Number Generation and Executive Functions in Parkinson's Disease: An Event-Related Brain Potential Study. *Journal of Parkinson's Disease*, *5*(3), 613-620.

*Muslimović, D., Post, B., Speelman, J. D., & Schmand, B. (2007). Motor procedural learning in Parkinson's disease. *Brain*, *130*(11), 2887-2897.

Nagano-Saito, A., Habak, C., Mejía-Constaín, B., Degroot, C., Monetta, L., Jubault, T., … Monchi, O. (2014). Effect of mild cognitive impairment on the patterns of neural activity in early Parkinson's disease. *Neurobiology of Aging*, *35*(1), 223-231.

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

*Nichelli, P., Appollonio, I., Clark, K., & Grafman, J. (1994). Word Frequency Monitoring in Parkinson Disease: An Analysis of Accuracy and Precision. *Cognitive and Behavioral Neurology*, *7*(4), 289-294.

*Nojszewska, M., Pilczuk, B., Zakrzewska-Pniewska, B., & Rowinska-Marcinska, K. (2009). The auditory system involvement in Parkinson disease: electrophysiological and neuropsychological correlations. *Journal of Clinical Neurophysiology*, *26*(6), 430-437.

Nyhus, E., & Barceló, F. (2009). The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: a critical update. *Brain and Cognition*, *71*, 437-451.

*Osternack Pinto, K. (2005). *Análise comparativa das funções neuropsicológicas de portadores de doença de Parkinson em estágios inicial e avançado: uma determinação de padrões para diagnóstico em população brasileira* (Unpublished doctoral dissertation). Universidade de São Paulo, São Paulo.

*Paolo, A. M., Tröster, A. I., Axelrod, B. N., & Koller, W. C. (1995). Construct validity of the WCST in normal elderly and persons with Parkinson's disease. *Archives of Clinical Neuropsychology*, *10*(5), 463-473.

Pascual-Sedano, B., Kulisevsky, J., Barbanoj, M., Garcia-Sanchez, C., Campolongo, A., Gironell, A., ... & Gich, I. (2008). Levodopa and executive performance in Parkinson's disease: a randomized study. *Journal of the International Neuropsychological Society*, *14*(5), 832-841.

*Passamonti, L., Salsone, M., Toschi, N., Cerasa, A., Giannelli, M., Chiriaco, C., ... & Quattrone, A. (2013). Dopamine-transporter levels drive striatal responses to apomorphine in Parkinson's disease. *Brain and Behavior*, *3*(3), 249-262.

*Pavlova, R., Mehrabian, S., Petrova, M., Skelina, S., Mihova, K., Jordanova, A., ... & Traykov, L. (2014). Cognitive, Neuropsychiatric, and Motor Features Associated With Apolipoprotein E∈ 4 Allele in a Sample of Bulgarian Patients With Late-Onset Parkinson's Disease. *American Journal of Alzheimer's Disease & Other Dementias*, *29*(7), 614-619.

*Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 275-288.

*Pellicano, C., Assogna, F., Cravello, L., Langella, R., Caltagirone, C., Spalletta, G., & Pontieri, F. E. (2015). Neuropsychiatric and cognitive symptoms and body side of onset of

parkinsonism in unmedicated Parkinson's disease patients. *Parkinsonism & Related Disorders*, *21*(9), 1096-1100.

*Pellicano, C., Assogna, F., Piras, F., Caltagirone, C., Pontieri, F. E., & Spalletta, G. (2012). Regional cortical thickness and cognitive functions in non-demented Parkinson's disease patients: a pilot study. *European Journal of Neurology*, *19*(1), 172-175.

*Perfetti, B., Varanese, S., Mercuri, P., Mancino, E., Saggino, A., & Onofrj, M. (2010). Behavioural assessment of dysexecutive syndrome in Parkinson's disease without dementia: a comparison with other clinical executive tasks. *Parkinsonism & Related Disorders*, *16*(1), 46-50.

*Péron, J., Biseul, I., Leray, E., Vicente, S., Le Jeune, F., Drapier, S., ... & Vérin, M. (2010a). Subthalamic nucleus stimulation affects fear and sadness recognition in Parkinson's disease. *Neuropsychology*, *24*(1), 1-8.

*Péron, J., Grandjean, D., Drapier, S., & Vérin, M. (2014). Effect of dopamine therapy on nonverbal affect burst recognition in Parkinson's disease. *PloS one*, *9*(3): e90092.

*Péron, J., Grandjean, D., Le Jeune, F., Sauleau, P., Haegelen, C., Drapier, D., ... & Vérin, M. (2010b). Recognition of emotional prosody is altered after subthalamic nucleus deep brain stimulation in Parkinson's disease. *Neuropsychologia*, *48*(4), 1053-1062.

*Péron, J., Le Jeune, F., Haegelen, C., Dondaine, T., Drapier, D., Sauleau, P., ... & Vérin, M. (2010c). Subthalamic nucleus stimulation affects theory of mind network: a PET study in Parkinson's disease. *PLoS One*, *5*(3): e9919.

*Perretta, J. G., Pari, G., & Beninger, R. J. (2005). Effects of Parkinson disease on two putative nondeclarative learning tasks: probabilistic classification and gambling. *Cognitive and Behavioral Neurology*, *18*(4), 185-192.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*, *295*(6), 676-680.

*Petrova, M., Raycheva, M., & Traykov, L. (2012). Cognitive profile of the earliest stage of dementia in Parkinson's disease. *American Journal of Alzheimer's Disease & Other Dementias*, *27*(8), 614-619.

*Petrova, M., Raycheva, M., Zhelev, Y., & Traykov, L. (2010). Executive functions deficit in Parkinson's disease with amnestic mild cognitive impairment. *American Journal of Alzheimer's Disease & Other Dementias*, *25*(5), 455-460.

*Pillon, B., Ertle, S., Deweer, B., Sarazin, M., Agid, Y., & Dubois, B. (1996). Memory for spatial location is affected in Parkinson's disease. *Neuropsychologia*, *34*(1), 77-85.

*Pirogovsky-Turk, E., Moore, R. C., Filoteo, J. V., Litvan, I., Song, D. D., Lessig, S. L., & Schiehser, D. M. (2017). Neuropsychiatric Predictors of Cognitive Decline in Parkinson Disease: A Longitudinal Study. *The American Journal of Geriatric Psychiatry*, *25*(3), 279-289.

*Poletti, M., Frosini, D., Pagni, C., Baldacci, F., Giuntini, M., Mazzucchi, S., ... & Bonuccelli, U. (2013). The relationship between motor symptom lateralization and cognitive performance in newly diagnosed drug-naïve patients with Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *35*(2), 124-131.

*Possin, K. L. (2007). *Visuospatial and visual object cognition in early Parkinson's disease* (Unpublished doctoral dissertation). University of California, San Diego and San Diego State University, San Diego.

*Pozzi, D., Petracchi, M., Sabe, L., Dancygier, G., García, H., & Starkstein, S. (1994). Quantified electroencephalographic changes in Parkinson's disease with and without dementia. *European Journal of Neurology*, *1*(2), 147-152.

*Price, A. L. (2005). Cortico-striatal contributions to category learning: dissociating the verbal and implicit systems. *Behavioral Neuroscience*, *119*(6), 1438-1447.

*Price, A. L. (2006). Explicit category learning in Parkinson's disease: Deficits related to impaired rule generation and selection processes. *Neuropsychology*, *20*(2), 249-257.

*Price, A. L. (2010). Contextual cues remediate problem solving: Deficits in Parkinson's disease. *Journal of the International Neuropsychological Society*, *16*(1), 70-76.

*Price, A., & Shin, J. C. (2009). The impact of Parkinson's disease on sequence learning: perceptual pattern learning and executive function. *Brain and Cognition*, *69*(2), 252-261.

*Puertas-Martín, V., Villarejo-Galende, A., Fernández-Guinea, S., Romero, J. P., Louis, E. D., & Benito-León, J. (2016). A Comparison Study of Cognitive and Neuropsychiatric Features of Essential Tremor and Parkinson's Disease. *Tremor and Other Hyperkinetic Movements*, *6:* 431.

*Ravizza, S. M., & Ciranni, M. A. (2002). Contributions of the prefrontal cortex and basal ganglia to set shifting. *Journal of Cognitive Neuroscience*, *14*(3), 472-483.

Ridderinkhof, K. R., Span, M. M., & Van Der Molen, M. W. (2002). Perseverative behavior and adaptive control in older adults: performance monitoring, rule induction, and set shifting. *Brain and Cognition*, *49*(3), 382-401.

Robbins, T. W., & Cools, R. (2014). Cognitive deficits in Parkinson's disease: a cognitive neuroscience perspective. *Movement Disorders*, *29*(5), 597–607.

Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L., & Treasure, J. (2007). A systematic review and meta-analysis of set-shifting ability in eating disorders. *Psychological Medicine*, *37*(08), 1075-1084.

*Roca, M., Manes, F., Chade, A., Gleichgerrcht, E., Gershanik, O., Arévalo, G. G., ... & Duncan, J. (2012). The relationship between executive functions and fluid intelligence in Parkinson's disease. *Psychological Medicine*, *42*(11), 2445-2452.

Romine, C. B., Lee, D., Wolfe, M. E., Homack, S., George, C., & Riccio, C. A. (2004). Wisconsin Card Sorting Test with children: a meta-analytic study of sensitivity and specificity. *Archives of Clinical Neuropsychology*, *19*(8), 1027-1041.

*Rosen, J. B., Brand, M., Polzer, C., Ebersbach, G., & Kalbe, E. (2013). Moral decision-making and theory of mind in patients with idiopathic Parkinson's disease. *Neuropsychology*, *27*(5), 562-572.

*Rosen, J. B., Rott, E., Ebersbach, G., & Kalbe, E. (2015). Altered moral decision-making in patients with idiopathic Parkinson's disease. *Parkinsonism & Related Disorders*, *21*(10), 1191-1199.

*Rouillard, M., Audiffren, M., Albinet, C., Ali Bahri, M., Garraux, G., & Collette, F. (2017). Contribution of four lifelong factors of cognitive reserve on late cognition in normal aging and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *39*(2), 142-162.

*Sagar, H. J., Sullivan, E. V., Cooper, J. A., & Jordan, N. (1991). Normal release from proactive interference in untreated patients with Parkinson's disease. *Neuropsychologia*, *29*(11), 1033-1044.

*Sánchez, J. L., Rodríguez, M., & Carro, J. (2002). Influence of cognitive reserve on neuropsychological functioning in Parkinson's disease. *Acta Neuropsychiatrica*, *14*(5), 207-215.

*Schmidt, N. (2014). *Neuropsychologische und strukturelle Korrelate der Riechleistung bei Patienten mit Morbus Parkinson* (Unpublished doctoral dissertation). Christian-Albrechts Universität zu Kiel, Kiel.

Schretlen, D. J. (2010). *Modified Wisconsin card sorting test (M-WCST).* Lutz, FL: Psychological Assessment Resources.

Seer, C., Lange, F., Georgiev, D., Jahanshahi, M., & Kopp, B. (2016). Event-related potentials and cognition in Parkinson's disease: An integrative review. *Neuroscience & Biobehavioral Reviews, 71*, 691-714.

Shin, N. Y., Lee, T. Y., Kim, E., & Kwon, J. S. (2014). Cognitive functioning in obsessive-compulsive disorder: a meta-analysis. *Psychological Medicine*, *44*(06), 1121-1130.

*Smith, J. G., & McDowall, J. (2006a). The implicit sequence learning deficit in patients with Parkinson's disease: a matter of impaired sequence integration? *Neuropsychologia, 44*(2), 275-288.

*Smith, J. G., & McDowall, J. (2006b). When artificial grammar acquisition in Parkinson's disease is impaired: The case of learning via trial-by-trial feedback. *Brain Research, 1067*(1), 216-228.

*Smith, J. G., & Mcdowall, J. (2011). Dissociating sequence learning performance in Parkinson's disease: visuomotor sequence acquisition and pattern judgment on a serial reaction time task. *Acta Neurobiologiae Experimentalis*, *71*, 359-380.

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin*, *139*(1), 81-132.

*Stamenović, J. D., Đurić, S. S., & Rančić, M. L. (2003). Cognitive dysfunctions in early stages of the Parkinson's disease. *Medicus*, *4*(2), 35-38.

*Stefanova, E. D., Kostic, V. S., Ziropadja, L. J., Ocic, G. G., & Markovic, M. (2001). Declarative memory in early Parkinsons disease: Serial position learning effects. *Journal of Clinical and Experimental Neuropsychology*, *23*(5), 581-591.

*Strohmaier, T. M. (2016). *Der Interlocking Finger Test als Screening-Instrument zur Detektion visuell-räumlicher Defizite und einer dementiellen Entwicklung bei Patienten mit Morbus Parkinson: eine Pilotstudie* (Unpublished doctoral dissertation). Christian-Albrechts Universität zu Kiel, Kiel.

*Taylor, A. E., Saint-Cyr, J. A., & Lang, A. E. (1986). Frontal lobe dysfunction in Parkinson's disease: the cortical focus of neostriatal outflow. *Brain*, *109*(5), 845-883.

*Tomer, R., Fisher, T., Giladi, N., & Aharon-Peretz, J. (2002). Dissociation between spontaneous and reactive flexibility in early Parkinson's disease. *Cognitive and Behavioral Neurology*, *15*(2), 106-112.

*Torralva, T., Laffaye, T., Báez, S., Gleichgerrcht, E., Bruno, D., Chade, A., ... & Roca, M. (2015). Verbal Fluency as a Rapid Screening Test for Cognitive Impairment in Early Parkinson's Disease. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *27*(3), 244-247.

*Tröster, A. I., Fields, J. A., Paolo, A. M., & Koller, W. C. (2006). Absence of the apolipoprotein E ε4 allele is associated with working memory impairment in Parkinson's disease. *Journal of the Neurological Sciences*, *248*(1), 62-67.

*Tröster, A. I., Stalp, L. D., Paolo, A. M., Fields, J. A., & Koller, W. C. (1995). Neuropsychological impairment in Parkinson's disease with and without depression. *Archives of Neurology*, *52*(12), 1164-1169.

*Vance, K. T. (1990). *Problem-solving abilities in Parkinson's disease* (Unpublished doctoral dissertation). The University of Arizona, Arizona.

*Venneri, A., Nichelli, P., Modonesi, G., Molinari, M. A., Russo, R., & Sardini, C. (1997). Impairment in dating and retrieving remote events in patients with early Parkinson9s disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *62*(4), 410-413.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428-443.

*Vicente, S., Péron, J., Biseul, I., Ory, S., Philippot, P., Drapier, S., ... & Vérin, M. (2011). Subjective emotional experience at different stages of Parkinson's disease. *Journal of the Neurological Sciences*, *310*(1), 241-247.

Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*, 14(1), 135.

*Werheid, K., Koch, I., Reichert, K., & Brass, M. (2007). Impaired self-initiated task preparation during task switching in Parkinson's disease. *Neuropsychologia*, *45*(2), 273-281.

*Wild, L. B., de Lima, D. B., Balardin, J. B., Rizzi, L., Giacobbo, B. L., Oliveira, H. B., ... & Bromberg, E. (2013). Characterization of cognitive and motor performance during dual-tasking in healthy older adults and patients with Parkinson's disease. *Journal of Neurology*, *260*(2), 580-589.

*Willemssen, R., Müller, T., Schwarz, M., Falkenstein, M., & Beste, C. (2009). Response monitoring in de novo patients with Parkinson's disease. *PloS one*, *4*(3): e4898.

*Willemssen, R., Müller, T., Schwarz, M., Hohnsbein, J., & Falkenstein, M. (2008). Error processing in patients with Parkinson's disease: the influence of medication state. *Journal of Neural Transmission*, *115*(3), 461-468.

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.

*Witt, K., Daniels, C., Daniel, V., Schmitt-Eliassen, J., Volkmann, J., & Deuschl, G. (2006a). Patients with Parkinson's disease learn to control complex systems—an indication for intact implicit cognitive skill learning. *Neuropsychologia*, *44*(12), 2445-2451.

*Witt, K., Daniels, C., Schmitt-Eliassen, J., Kernbichler, J., Rehm, S., Volkmann, J., & Deuschl, G. (2006b). The impact of normal aging and Parkinson's disease on response preparation in task-switching behavior. *Brain Research*, *1114*(1), 173-182.

*Witt, K., Nuhsman, A., & Deuschl, G. (2002). Dissociation of habit-learning in Parkinson's and cerebellar disease. *Journal of Cognitive Neuroscience*, *14*(3), 493-499.

*Woods, S. P., & Tröster, A. I. (2003). Prodromal frontal/executive dysfunction predicts incident dementia in Parkinson's disease. *Journal of the International Neuropsychological Society*, *9*(1), 17-24.

Wuensch, K. L. Using SPSS to obtain a confidence interval for Cohen's d. (2012). http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.pdf Accessed 30.03.17.

*Yu, R. L., Wu, R. M., Chiu, M. J., Tai, C. H., Lin, C. H., & Hua, M. S. (2012). Advanced Theory of Mind in patients at early stage of Parkinson’s disease. *Parkinsonism & Related Disorders*, *18*(1), 21-24.

*Yu, R. L., Wu, R. M., Tai, C. H., Lin, C. H., & Hua, M. S. (2010b). Feeling-of-knowing in episodic memory in patients with Parkinson's disease with various motor symptoms. *Movement Disorders*, *25*(8), 1034-1039.

*Yu, R. L., Wu, R. M., Tai, C. H., Lin, C. H., Cheng, T. W., & Hua, M. S. (2012a). Neuropsychological profile in patients with early stage of Parkinson's disease in Taiwan. *Parkinsonism & Related Disorders*, *18*(10), 1067-1072.

*Zeng, X. H., Hirata, K., Tanaka, H., Hozumi, A., & Yamazaki, K. (2002). Insufficient processing resources in Parkinson's disease: evaluation using multimodal event-related potentials paradigm. *Brain Topography, 14*(4), 299-311.

Zgaljardic, D. J., Borod, J. C., Foldi, N. S., & Mattis, P. (2003). A review of the cognitive and behavioral sequelae of Parkinson's disease: relationship to frontostriatal circuitry. *Cognitive and Behavioral Neurology, 16*(4), 193-210.

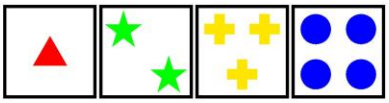
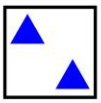
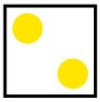
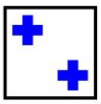
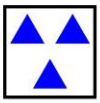
## Figure Captions

*Figure 1*. The first thirteen trials completed by a hypothetical examinee on the Modified Wisconsin Card Sorting Test (M-WCST; Schretlen, 2010). On this version of the test, a category is considered completed after six consecutive sorts according to the correct rule. In contrast, the widely disseminated Wisconsin Card Sorting Test version by Heaton and colleagues (1993) requires ten consecutive correct responses. Here, the hypothetical examinee needs eight trials to complete the first category (trials to criterion = 8). Over the first thirteen trials, the examinee commits five errors, with two of them being perseverative (i.e., repetitions of a rule that whose application has resulted in negative feedback on the previous trial) and the other three being non-perseverative errors. The individual also commits four perseverative responses (i.e., sorts according to the previously correct rule). Note that the number of perseverative responses and the number of trials to reach the first criterion are not scored within the M-WCST, but within the Heaton et al. version of the test. The same applies to the percentage of conceptual level responses (i.e., consecutive correct responses occurring in runs of three) and the number of failures to maintain set (i.e., errors that are made after five consecutive correct responses but before the category is completed). C = color, S = shape, N = number.

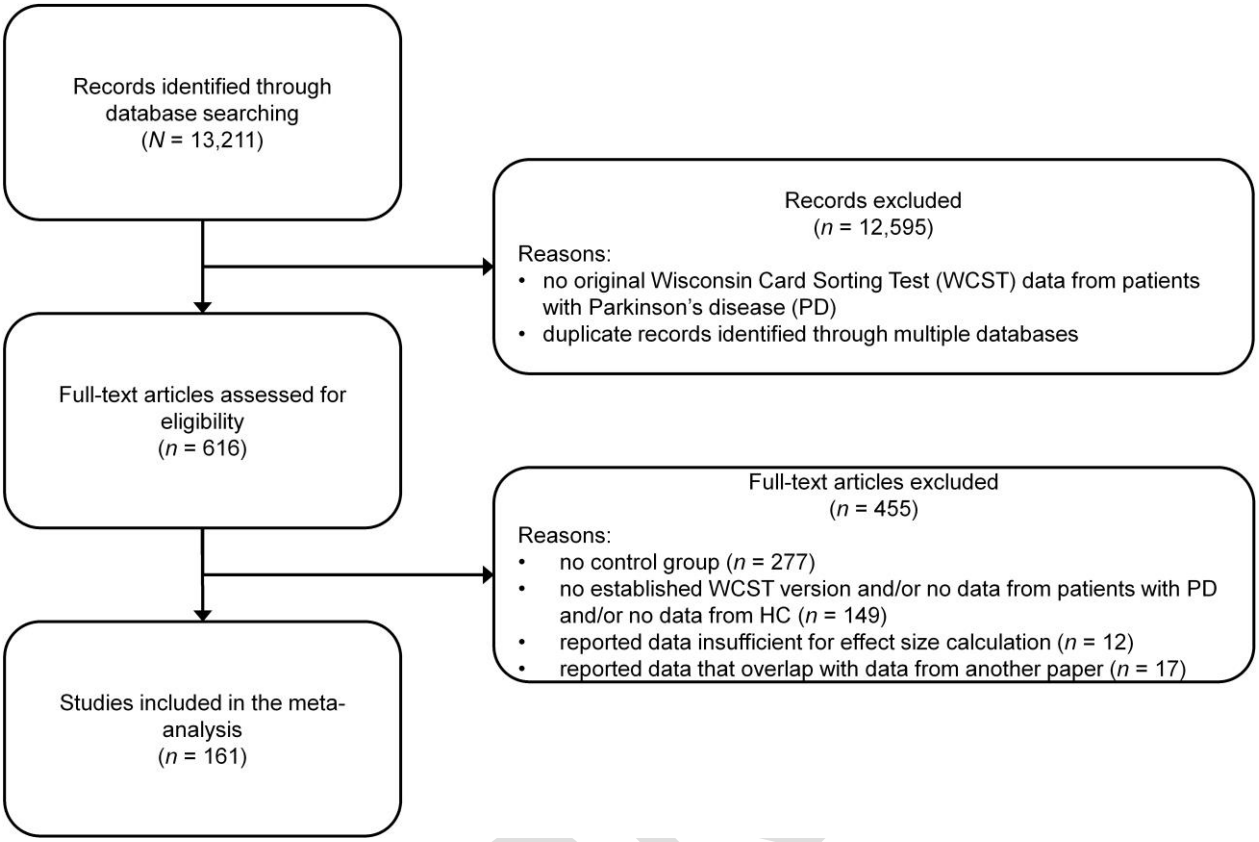
*Figure 2*. Flow chart depicting the selection of articles for our meta-analysis.

*Figure 3*. Mean effect sizes for the difference in the number of WCST categories completed by patients with Parkinson's disease and healthy control participants as a function of disease duration, patients' scores on the motor scale of the Unified Parkinson's Disease Rating Scale (UPDRS), and patients' medication status. The vertical line reflects the mean effect size from our random-effects meta-analyses ($d$ = 0.74) for comparison.

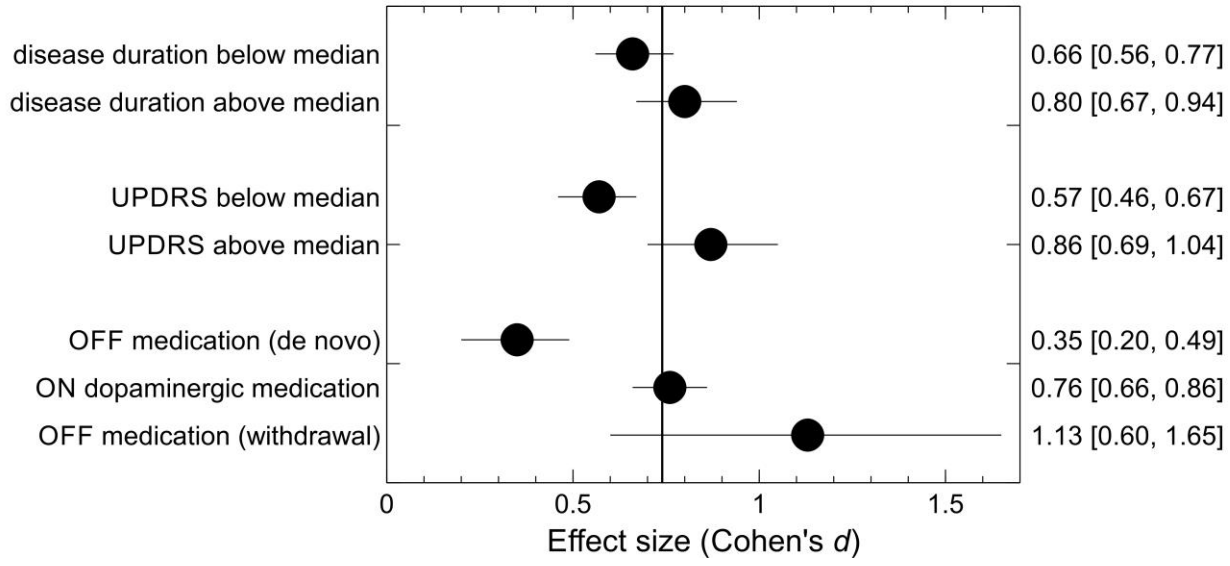
*Figure 4.* Mean effect sizes for the difference in the number of WCST categories completed by patients with Parkinson's disease and healthy control participants as a function of patients' scores on the motor scale of the Unified Parkinson's Disease Rating Scale (UPDRS). Within each quartile, studies are listed in ascending order according to their sample sizes. The area of the circles is proportional to the studies' sample sizes.

| trial | correct rule |  | applied rule | sequence correct | feedback | scoring |
|---|---|---|---|---|---|---|
| 1 | C |  | C | 1 | correct! | |
| 2 | C |  | N | - | wrong! | non-perseverative error |
| 3 | C |  | C | 1 | correct! | |
| 4 | C |  | C | 2 | correct! | |
| 5 | C |  | C | 3 | correct! | |
| 6 | C |  | C | 4 | correct! | |
| 7 | C |  | C | 5 | correct! | |
| 8 | C |  | C | 6 | correct! | category completed |
| 9 | S |  | C | - | wrong! | non-perseverative error/perseverative response |
| 10 | S |  | C | - | wrong! | perseverative error/perseverative response |
| 11 | S |  | S | 1 | correct! | |
| 12 | S |  | C | - | wrong! | non-perseverative error/perseverative response |
| 13 | S |  | C | - | wrong! | perseverative error/perseverative response |

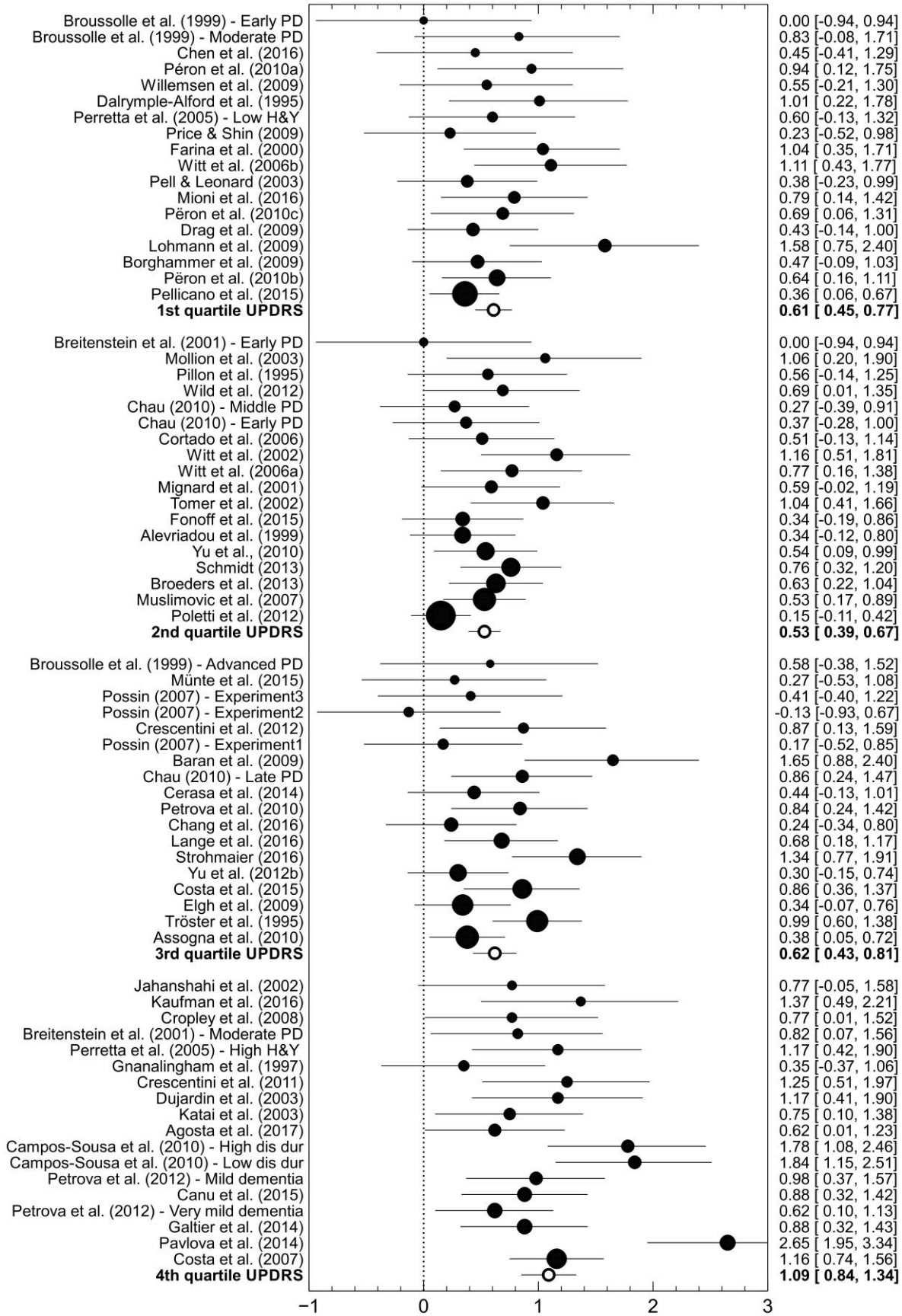
1330


Records identified through database searching (N = 13,211)
Records excluded (n = 12,595)
Reasons:
• no original Wisconsin Card Sorting Test (WCST) data from patients with Parkinson's disease (PD)
• duplicate records identified through multiple databases
Full-text articles assessed for eligibility (n = 616)
Full-text articles excluded (n = 455)
Reasons:
• no control group (n = 277)
• no established WCST version and/or no data from patients with PD and/or no data from HC (n = 149)
• reported data insufficient for effect size calculation (n = 12)
• reported data that overlap with data from another paper (n = 17)
Studies included in the meta-analysis (n = 161)

1331


disease duration below median
disease duration above median
UPDRS below median
UPDRS above median
OFF medication (de novo)
ON dopaminergic medication
OFF medication (withdrawal)
0.66 [0.56, 0.77]
0.80 [0.67, 0.94]
0.57 [0.46, 0.67]
0.86 [0.69, 1.04]
0.35 [0.20, 0.49]
0.76 [0.66, 0.86]
1.13 [0.60, 1.65]
0
0.5
1
1.5
Effect size (Cohen's d)

1332


Broussolle et al. (1999) - Early PD 0.00 [-0.94, 0.94]
Broussolle et al. (1999) - Moderate PD 0.83 [-0.08, 1.71]
Chen et al. (2016) 0.45 [-0.41, 1.29]
Péron et al. (2010a) 0.94 [ 0.12, 1.75]
Willemsen et al. (2009) 0.55 [-0.21, 1.30]
Dalrymple-Alford et al. (1995) 1.01 [ 0.22, 1.78]
Perretta et al. (2005) - Low H&Y 0.60 [-0.13, 1.32]
Price & Shin (2009) 0.23 [-0.52, 0.98]
Farina et al. (2000) 1.04 [ 0.35, 1.71]
Witt et al. (2006b) 1.11 [ 0.43, 1.77]
Pell & Leonard (2003) 0.38 [-0.23, 0.99]
Mioni et al. (2016) 0.79 [ 0.14, 1.42]
Përon et al. (2010c) 0.69 [ 0.06, 1.31]
Drag et al. (2009) 0.43 [-0.14, 1.00]
Lohmann et al. (2009) 1.58 [ 0.75, 2.40]
Borghammer et al. (2009) 0.47 [-0.09, 1.03]
Përon et al. (2010b) 0.64 [ 0.16, 1.11]
Pellicano et al. (2015) 0.36 [ 0.06, 0.67]
1st quartile UPDRS 0.61 [ 0.45, 0.77]
Breitenstein et al. (2001) - Early PD 0.00 [-0.94, 0.94]
Mollion et al. (2003) 1.06 [ 0.20, 1.90]
Pillon et al. (1995) 0.56 [-0.14, 1.25]
Wild et al. (2012) 0.69 [ 0.01, 1.35]
Chau (2010) - Middle PD 0.27 [-0.39, 0.91]
Chau (2010) - Early PD 0.37 [-0.28, 1.00]
Cortado et al. (2006) 0.51 [-0.13, 1.14]
Witt et al. (2002) 1.16 [ 0.51, 1.81]
Witt et al. (2006a) 0.77 [ 0.16, 1.38]
Mignard et al. (2001) 0.59 [-0.02, 1.19]
Tomer et al. (2002) 1.04 [ 0.41, 1.66]
Fonoff et al. (2015) 0.34 [-0.19, 0.86]
Alevriadou et al. (1999) 0.34 [-0.12, 0.80]
Yu et al., (2010) 0.54 [ 0.09, 0.99]
Schmidt (2013) 0.76 [ 0.32, 1.20]
Broeders et al. (2013) 0.63 [ 0.22, 1.04]
Muslimovic et al. (2007) 0.53 [ 0.17, 0.89]
Poletti et al. (2012) 0.15 [-0.11, 0.42]
2nd quartile UPDRS 0.53 [ 0.39, 0.67]
Broussolle et al. (1999) - Advanced PD 0.58 [-0.38, 1.52]
Műnte et al. (2015) 0.27 [-0.53, 1.08]
Possin (2007) - Experiment3 0.41 [-0.40, 1.22]
Possin (2007) - Experiment2 -0.13 [-0.93, 0.67]
Crescentini et al. (2012) 0.87 [ 0.13, 1.59]
Possin (2007) - Experiment1 0.17 [-0.52, 0.85]
Baran et al. (2009) 1.65 [ 0.88, 2.40]
Chau (2010) - Late PD 0.86 [ 0.24, 1.47]
Cerasa et al. (2014) 0.44 [-0.13, 1.01]
Petrova et al. (2010) 0.84 [ 0.24, 1.42]
Chang et al. (2016) 0.24 [-0.34, 0.80]
Lange et al. (2016) 0.68 [ 0.18, 1.17]
Strohmaier (2016) 1.34 [ 0.77, 1.91]
Yu et al. (2012b) 0.30 [-0.15, 0.74]
Costa et al. (2015) 0.86 [ 0.36, 1.37]
Elgh et al. (2009) 0.34 [-0.07, 0.76]
Tröster et al. (1995) 0.99 [ 0.60, 1.38]
Assogna et al. (2010) 0.38 [ 0.05, 0.72]
3rd quartile UPDRS 0.62 [ 0.43, 0.81]
Jahanshahi et al. (2002) 0.77 [-0.05, 1.58]
Kaufman et al. (2016) 1.37 [ 0.49, 2.21]
Cropley et al. (2008) 0.77 [ 0.01, 1.52]
Breitenstein et al. (2001) - Moderate PD 0.82 [ 0.07, 1.56]
Perretta et al. (2005) - High H&Y 1.17 [ 0.42, 1.90]
Gnanalingham et al. (1997) 0.35 [-0.37, 1.06]
Crescentini et al. (2011) 1.25 [ 0.51, 1.97]
Dujardin et al. (2003) 1.17 [ 0.41, 1.90]
Katai et al. (2003) 0.75 [ 0.10, 1.38]
Agosta et al. (2017) 0.62 [ 0.01, 1.23]
Campos-Sousa et al. (2010) - High dis dur 1.78 [ 1.08, 2.46]
Campos-Sousa et al. (2010) - Low dis dur 1.84 [ 1.15, 2.51]
Petrova et al. (2012) - Mild dementia 0.98 [ 0.37, 1.57]
Canu et al. (2015) 0.88 [ 0.32, 1.42]
Petrova et al. (2012) - Very mild dementia 0.62 [ 0.10, 1.13]
Galtier et al. (2014) 0.88 [ 0.32, 1.43]
Pavlova et al. (2014) 2.65 [ 1.95, 3.34]
Costa et al. (2007) 1.16 [ 0.74, 1.56]
4th quartile UPDRS 1.09 [ 0.84, 1.34]
−1
0
1
2
3
effect size (Cohen's d)

1333

Table 1.

*Overview of the studies included in the meta-analysis of Wisconsin Card Sorting Test (WCST) performance in patients with Parkinson's Disease*

| Study | $N_{PD}$ | $N_{HC}$ | age | %F | $\Delta_{age}$ | $\Delta_{\%F}$ | dur | HY | UPDRS | MMSE | $\Delta_{MMSE}$ | med | dem | dep | $D_{categories}$ | $D_{selected\ pers.\ measure}$ | selected measure of perseveration |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Abo-El-Naga (2006) | 43 | 30 | 65.2 | 46.5 | 2.5 | -0.2 | 15.8 | 1.8 | 16.7 | 27.8 | -0.4 | M (ON) | E | NE | | 0.55 [0.07, 1.02] | Perseverations |
| Agosta et al. (2017) | 25 | 19 | 66.4 | 28.0 | 0.4 | -24.6 | - | 2.3 | 25.5 | 27.7 | -1.5 | ON | E | E | 0.62 [0.01, 1.23] | 0.61 [0, 1.22] | Perseverations |
| Alevriadou et al. (1999) | 37 | 37 | 62.1 | 37.8 | 1.9 | -24.4 | 110.4 | 2.6 | 16.7 | - | - | ON | E | NE | 0.34 [-0.12, 0.80] | 0.51 [0.04, 0.97] | Perseverative errors |
| Alonso Recio et al. (2013) | 23 | 18 | 65.5 | - | 0.4 | - | 78.1 | 1.5 | - | - | - | ON | E | NE | 0.55 [-0.08, 1.18] | | |
| Asahina et al. (1996) | 10 | 8 | 60.5 | 70.0 | -2.3 | 20.0 | 43.8 | 2.3 | - | 28.3 | -0.1 | OFF (M) | E | NE | 1.33 [0.27, 2.35] | 1.01 [0, 1.99] | Perseverative errors (n) |
| Assogna et al. (2010) | 70 | 70 | 62.2 | 52.0 | 0.1 | 0.0 | 58.8 | - | 20.1 | 27.9 | -1.2 | ON | E | NE | 0.38 [0.05, 0.72] | 0.41 [0.08, 0.75] | Perseverative errors |
| Azuma et al. (2003) | 69 | 37 | 68.9 | 38.0 | 0.8 | -21.0 | 68.4 | - | - | 28.4 | -0.6 | ON | E | NE | 0.65 [0.24, 1.06] | | |
| Baran et al. (2009) | 18 | 18 | 65.3 | 16.7 | 1.6 | -16.3 | - | 1.3 | 20.4 | 26.7 | -2.9 | ON | E | NE | 1.65 [0.88, 2.40] | 1.05 [0.34, 1.74] | Perseverative errors (%) |
| Beatty & Monson (1990) | 27 | 25 | 66.4 | 40.7 | 0.7 | -7.3 | 75.6 | 2.6 | - | - | - | M (ON) | NE | NE | 0.55 [-0.01, 1.10] | 0.32 [-0.23, 0.87] | Perseverative errors |
| Beatty et al. (1989) - Demented | 18 | 15 | 68.2 | - | 4.3 | - | 75.6 | - | - | 25.2 | -3.9 | M (ON) | NE | NE | 0.83 [0.11, 1.54] | 0.88 [0.16, 1.60] | Perseverative errors |
| Beatty et al. (1989) - Non-demented | 25 | 13 | 64.1 | - | -1.3 | - | 55.2 | - | - | 28.9 | -0.3 | M (ON) | E | NE | 0.94 [0.23, 1.64] | 0.58 [-0.11, 1.26] | Perseverative errors |
| Blonder et al. (1989) | 21 | 17 | 61.3 | 47.6 | -1.1 | -5.3 | 50.9 | 1.2 | - | - | - | ON | NE | NE | 1.02 [0.34, 1.70] | 0.69 [0.03, 1.35] | Perseverative errors |
| Bokura et al. (2005) | 13 | 14 | 71.0 | 38.5 | 0.0 | -11.5 | - | 2.9 | - | - | - | ON | E | NE | 1.41 [0.55, 2.25] | | |
| Borghammer et al. (2009) | 24 | 26 | 62.0 | 41.7 | 2.0 | -19.8 | 44.4 | 1.5 | 13.7 | 29.0 | 0.3 | ON | E | NE | 0.47 [-0.09, 1.03] | | |
| Brand et al. (2004) | 20 | 20 | 66.9 | 45.0 | 2.9 | 15.0 | 106.1 | 3.0 | - | 28.2 | - | ON | E | NE | 1.71 [0.97, 2.43] | 0.68 [0.04, 1.32] | Perseverations |
| Breitenstein et al. (2001) - Early PD | 6 | 16 | 68.3 | 33.3 | -0.3 | -16.7 | 16.2 | 2.0 | 17.5 | - | - | OFF (DN) | E | NE | 0 [-0.94, 0.94] | 0 [-0.94, 0.94] | Perseverative errors (n) |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Breitenstein et al. (2001) - Moderate PD | 14 | 16 | 72.6 | 35.7 | 4.0 | -14.3 | 59.1 | 2.1 | 27.5 | - | - | ON | E | NE | 0.82 [0.07, 1.56] | 0.68 [-0.06, 1.42] | Perseverative errors (n) |
| Broeders et al. (2013) | 59 | 40 | 62.5 | 45.8 | 1.1 | 0.8 | 17.5 | 1.7 | 16.0 | 27.9 | -1.0 | ON | E | NE | 0.63 [0.22, 1.04] | 0.29 [-0.11, 0.69] | Perseverative errors (n) |
| Broussolle et al. (1999) - Advanced PD | 8 | 10 | 57.5 | 44.4 | 4.6 | -15.6 | 148.8 | 1.9 | 20.6 | - | - | ON | E | NE | 0.58 [-0.38, 1.52] | 0.63 [-0.33, 1.58] | Perseverative errors (n) |
| Broussolle et al. (1999) - Early PD | 8 | 10 | 54.5 | 44.4 | 1.6 | -15.6 | 15.6 | 1.4 | 11.9 | - | - | OFF (M) | E | NE | 0 [-0.94, 0.94] | -0.28 [-1.21, 0.66] | Perseverative errors (n) |
| Broussolle et al. (1999) - Moderate PD | 11 | 10 | 55.7 | 44.4 | 2.8 | -15.6 | 86.4 | 1.4 | 12.5 | - | - | ON | E | NE | 0.83 [-0.08, 1.71] | 0.51 [-0.37, 1.37] | Perseverative errors (n) |
| Brown & Marsden (1988a) | 16 | 16 | 60.3 | 31.3 | 1.4 | - | 139.2 | 2.3 | - | - | - | ON | E | NE | 1.32 [0.54, 2.08] | 1.28 [0.51, 2.04] | Perseverative errors |
| Brown & Marsden (1988b) | 16 | 16 | 59.2 | - | 3.1 | - | 134.4 | 2.5 | - | - | - | ON | E | NE | 1.35 [0.57, 2.11] | 1.07 [0.32, 1.81] | Perseverative errors |
| Brown et al. (2002) - Experiment1 | 24 | 30 | 68.2 | 37.5 | 1.6 | -19.2 | - | 1.8 | - | - | - | ON | E | NE | 0.75 [0.19, 1.30] | 0.54 [-0.01, 1.09] | Perseverative errors |
| Brown et al. (2002) - Experiment2 | 17 | 16 | 70.9 | 29.4 | -0.1 | -26.9 | - | 2.3 | - | 28.2 | -0.9 | ON | E | NE | 0.61 [-0.10, 1.30] | 0.36 [-0.33, 1.05] | Perseverative errors |
| Caffarra et al. (2012) | 20 | 11 | 67.2 | 55.0 | -0.6 | 0.6 | - | - | - | 26.3 | -1.5 | - | NE | NE | 0.84 [0.07, 1.60] | | |
| Caltagirone et al. (2006) - Demented | 9 | 21 | 63.5 | - | 3.9 | - | 57.6 | 2.5 | - | - | - | M (OFF) | NE | NE | 1.98 [1.03, 2.91] | 1.60 [0.70, 2.47] | Perseverative errors |
| Caltagirone et al., (2006) – Non-demented | 15 | 21 | 60.0 | - | 0.4 | - | 43.2 | 2.5 | - | - | - | M (OFF) | E | NE | 0.74 [0.05, 1.42] | 0.75 [0.06, 1.43] | Perseverative errors |
| Campos-Sousa et al. (2010) - High dis dur | 21 | 25 | 59.7 | 52.4 | 0.6 | -15.6 | 78.24 | - | 35.2 | - | - | ON | E | NE | 1.78 [1.08, 2.46] | 0.91 [0.30, 1.52] | Perseverative responses |
| Campos-Sousa et al. (2010) - Low dis dur | 23 | 25 | 63.2 | 52.2 | 4.1 | -15.8 | 21.0 | - | 28.3 | - | - | ON | E | NE | 1.84 [1.15, 2.51] | 1.19 [0.57, 1.81] | Perseverative responses |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Canavan et al. (1989) | 19 | 10 | 57.9 | 31.6 | -2.4 | -18.4 | 34.8 | - | - | - | - | M (OFF) | NE | NE | | 0.92 [0.11, 1.71] | Perseverative errors |
| Canu et al. (2015) | 23 | 35 | 66.9 | 30.0 | -0.8 | -13 | - | 2.3 | 25.4 | 27.7 | -1.4 | ON | E | E | 0.88 [0.32, 1.42] | 0.63 [0.09, 1.17] | Perseverations |
| Cerasa et al. (2014) | 24 | 24 | 58.7 | 8.3 | -1.7 | -8.4 | 58.8 | - | 19.9 | 27.8 | -1.0 | ON | E | NE | 0.44 [-0.13, 1.01] | 0.20 [-0.37, 0.76] | Perseverative errors |
| Chang et al. (2016) | 35 | 18 | 66.9 | 34.3 | -1.4 | -21.3 | 8.1 | 1.9 | 18.9 | 27.6 | -0.8 | ON | E | NE | 0.24 [-0.34, 0.80] | 1.75 [1.08, 2.41] | Perseverative errors (%) |
| Chau (2010) - Early PD | 18 | 20 | 62.6 | 39.0 | -3.0 | 4.0 | 9.6 | 1.0 | 17.7 | 28.7 | -0.9 | ON | E | NE | 0.37 [-0.28, 1.00] | | |
| Chau (2010) - Late PD | 25 | 20 | 64.5 | 25.0 | -1.1 | -10.0 | 91.2 | 2.0 | 22.6 | 28.4 | -1.2 | ON | E | NE | 0.86 [0.24, 1.47] | | |
| Chau (2010) - Middle PD | 17 | 20 | 64.4 | 24.0 | -1.2 | -11.0 | 44.4 | 1.0 | 15.8 | 28.8 | -0.8 | ON | E | NE | 0.27 [-0.39, 0.91] | | |
| Chen et al. (2006) | 27 | 27 | 63.3 | 37.0 | -0.2 | -14.9 | 40.1 | 2.0 | - | - | - | ON | E | NE | 0.19 [-0.35, 0.72] | -0.12 [-0.66, 0.41] | Perseverative errors (%) |
| Chen et al. (2016) | 10 | 12 | 63.7 | 70.0 | 1.1 | 3.3 | - | 1.6 | 9.2 | - | - | ON | E | NE | 0.45 [-0.41, 1.29] | | |
| Clark (2014) | 27 | 23 | 64.5 | 55.6 | 0.2 | -0.9 | 67.2 | 2.2 | 30.1 | 28.7 | 0.0 | OFF | E | NE | | | |
| Cohn et al. (2016) | 15 | 13 | 59.1 | 40.0 | 5.2 | 1.5 | 74.4 | 1.5 | 18.7 | - | - | M (ON) | E | NE | | -0.15 [-0.89, 0.60] | Perseverative errors |
| Cooper et al. (1991) | 60 | 37 | 59.8 | 48.3 | 0.2 | 2.4 | - | - | - | - | - | OFF (DN) | E | NE | 0.26 [-0.16, 0.67] | 0.36 [-0.05, 0.78] | Perseverative errors (%) |
| Cortado et al. (2006) | 17 | 23 | 67.7 | 23.5 | -3.8 | -15.6 | 94.3 | 2.6 | 18.9 | 28.6 | -0.8 | ON | NE | NE | 0.51 [-0.13, 1.14] | 0.50 [-0.14, 1.14] | Perseverative errors |
| Costa et al (2007) | 54 | 53 | 63.2 | 31.5 | 0.3 | -10.0 | 86.2 | 2.3 | 24.4 | 28.3 | -0.3 | ON | E | NE | 1.16 [0.74, 1.56] | 0.71 [0.32, 1.10] | Perseverative errors |
| Costa et al. (2015) | 81 | 20 | 64.9 | 44.4 | -1.1 | -40.6 | 78.0 | - | 21.3 | 28.9 | -0.5 | ON | NE | NE | 0.86 [0.36, 1.37] | 0.52 [0.02, 1.01] | Perseverative errors |
| Crescentini et al. (2011) | 19 | 16 | 66.7 | 26.3 | 1.1 | 1.3 | 76.8 | 2.1 | 26.2 | 29.0 | -0.2 | ON | E | NE | 1.25 [0.51, 1.97] | 0.68 [-0.01, 1.36] | Perseverative errors |
| Crescentini et al. (2012) | 16 | 16 | 63.6 | 50.0 | 2.0 | -12.5 | 72.0 | 1.9 | 22.9 | 28.8 | -0.5 | ON | E | NE | 0.87 [0.13, 1.59] | 0.41 [-0.30, 1.10] | Perseverative errors |
| Cropley et al. (2008) | 15 | 14 | 62.1 | 40.0 | 0.5 | -2.9 | 140.4 | 3.0 | 41.9 | 29.1 | -0.2 | OFF | E | NE | 0.77 [0.01, 1.52] | 0.44 [-0.30, 1.18] | Perseverative Responses |
| Dalrymple-Alford et al. (1994) | 7 | 7 | 65.6 | - | 3.2 | - | 52.8 | 2.1 | - | - | - | ON | E | NE | 0.67 [-0.42, 1.74] | 0 [-1.06, 1.06] | Perseverative errors (%) |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dalrymple-Alford et al. (1995) | 20 | 11 | 65.7 | 50.0 | 4.1 | 4.6 | 20.5 | 1.6 | 11.8 | - | - | M (OFF) | E | NE | 1.01 [0.22, 1.78] | 0.89 [0.11, 1.65] | Perseverative errors |
| Davidson et al. (2006) - Experiment1 | 19 | 23 | 67.1 | - | -0.4 | - | 69.5 | - | - | 29.3 | 0.0 | M (U) | NE | NE | | | |
| Davidson et al. (2006) - Experiment2 | 16 | 16 | 66.6 | - | -0.9 | - | 73.6 | - | - | 29.1 | 0.6 | ON | NE | NE | 0.45 [-0.17, 1.06] | -0.38 [-1.08, 0.32] | Perseverative errors |
| Davidson et al. (2013) | 18 | 23 | 71.0 | 33.3 | -1.0 | -5.8 | 120.0 | 2.0 | - | 27.8 | -0.1 | ON | E | NE | 0.67 [0.03, 1.30] | 0.65 [0.01, 1.28] | Perseverative errors (%) |
| Diaz-Santos et al. (2015) | 27 | 25 | 64.2 | - | -0.2 | - | 64.8 | 2.0 | - | 28.8 | 0.0 | ON | E | NE | | 0.43 [-0.12, 0.98] | Perseverative errors |
| Doyon et al. (1996) | 15 | 15 | 58.5 | 40.0 | 1.7 | 0.0 | 121.2 | 1.8 | - | 28.4 | -0.7 | M (ON) | E | NE | 0.28 [-0.44, 1.00] | 0.09 [-0.63, 0.81] | Perseverative errors (n) |
| Drag et al. (2009) | 24 | 24 | 69.0 | - | 0.4 | - | 53.0 | 1.9 | 14.4 | - | - | OFF | E | NE | 0.43 [-0.14, 1.00] | | |
| Dubois et al. (1988) | 33 | 20 | 60.4 | - | -2.6 | - | 104.4 | 2.6 | - | - | - | ON | NE | NE | | | |
| Dubois et al. (1990) - Early onset | 11 | 11 | 44.0 | - | 0.1 | - | 46.8 | 2.0 | - | - | - | ON | NE | NE | | | |
| Dubois et al. (1990) - Late onset | 11 | 11 | 72.7 | - | -0.8 | - | 43.2 | 2.5 | - | - | - | ON | NE | NE | | | |
| Dujardin et al. (2001) | 24 | 12 | 64.7 | 50.0 | 5.4 | 0.0 | 88.5 | 2.2 | 27.2 | - | - | M (ON) | E | NE | | 0.79 [0.07, 1.51] | Perseverative errors (%) |
| Dujardin et al. (2003) | 24 | 12 | 66.5 | 50.0 | 1.1 | -8.3 | 93.5 | - | 31.2 | 28.9 | -1.0 | ON | E | NE | 1.17 [0.41, 1.90] | 0.89 [0.16, 1.61] | Perseverative errors |
| Ebmeier et al. (1992) | 14 | 16 | 69.0 | 43.8 | 2.0 | 0.0 | 111.6 | 2.4 | - | - | - | ON | E | NE | 0.79 [0.04, 1.53] | 0.25 [-0.47, 0.97] | Perseverative errors (%) |
| Ekman et al. (2012) | 77 | 24 | 67.6 | 40.0 | -0.3 | -10.0 | 0.0 | - | 24.3 | 29.1 | -0.1 | OFF (DN) | E | NE | | -0.09 [-0.55, 0.37] | Persevere |
| Elgh et al. (2009) | 88 | 30 | 68.1 | 79.6 | -0.1 | 32.9 | 0.0 | - | 23.8 | 28.7 | -0.4 | M (OFF) | E | NE | 0.34 [-0.07, 0.76] | 0.27 [-0.14, 0.69] | Perseverative errors |
| Euteneuer et al. (2009) | 21 | 23 | 67.6 | 66.7 | 3.2 | 18.9 | 85.7 | 2.3 | 17.7 | 29.0 | -0.7 | ON | E | NE | | 0.38 [-0.22, 0.98] | Perseverations |
| Fales et al. (2006) | 21 | 25 | 66.9 | 65.0 | -1.9 | 17.0 | 69.6 | 2.0 | - | 28.8 | -0.2 | ON | E | NE | 0.55 [-0.04, 1.14] | -0.05 [-0.63, 0.53] | Perseverative errors |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fama et al. (2000) | 20 | 38 | 63.1 | - | -2.2 | - | 70.8 | - | - | 27.4 | -1.6 | ON | NE | NE | 0.80 [0.23, 1.36] | 0.48 [-0.07, 1.03] | Perseverative responses |
| Farina et al. (2000) | 20 | 18 | 57.9 | 35.0 | 1.3 | -9.4 | 28.0 | 1.5 | 9.1 | 27.8 | -1.3 | M (B) | E | NE | 1.04 [0.35, 1.71] | 1.03 [0.34, 1.70] | Perseverative errors |
| Filoteo et al. (2005) | 19 | 19 | 67.4 | 57.9 | 0.6 | -5.3 | 91.2 | 1.7 | - | - | - | ON | E | NE | -0.17 [-0.69, 0.36] | -0.19 [-0.83, 0.45] | Perseverative errors |
| Flensborg Damholdt et al. (2012) | 71 | 30 | 69.4 | - | 1.3 | - | 84.8 | - | - | 27.7 | -1.4 | ON | E | NE | 1.09 [0.63, 1.54] | 0.78 [0.34, 1.22] | Perseverative errors |
| Fonoff et al. (2015) | 28 | 28 | 59.3 | 42.9 | 0.0 | 12.2 | 159.6 | 2.8 | 16.2 | 28.4 | 0.2 | ON | NE | NE | 0.34 [-0.19, 0.86] | 0.68 [0.14, 1.22] | Perseverative errors |
| Galtier et al. (2014) | 43 | 20 | 59.2 | 44.2 | -1.7 | -10.8 | 99.6 | 2.3 | 28.5 | 27.6 | -0.8 | - | E | NE | 0.88 [0.32, 1.43] | | |
| Gasparini et al. (2001) | 15 | 15 | 66.6 | 46.7 | 1.0 | -6.6 | 86.4 | 2.5 | - | - | - | OFF | NE | NE | 1.58 [0.74, 2.40] | 1.82 [0.95, 2.67] | Perseverative errors |
| Gauggel et al. (2004) | 31 | 28 | 57.8 | 46.9 | 1.1 | 1.7 | 108.2 | 2.6 | - | - | - | ON | E | NE | | 0.57 [0.04, 1.09] | Perseverative errors (%) |
| Gawrys et al. (2008) | 19 | 21 | 57.0 | - | 1.3 | - | 53.4 | 1.9 | - | 29.2 | -0.3 | ON | E | NE | 0.65 [0.01, 1.28] | 0.95 [0.29, 1.60] | Perseverative errors |
| Gawrys et al. (2014) | 30 | 18 | 56.0 | 56.7 | -1.1 | 1.1 | 81.0 | 2.0 | - | 28.9 | -0.4 | M (ON) | E | NE | 1.53 [0.86, 2.18] | 1.25 [0.61, 1.89] | Perseverations |
| Gnanalingham et al. (1997) | 12 | 21 | 72.6 | 33.0 | -0.7 | -7.0 | 110.4 | - | 29.5 | 24.1 | -5.0 | M (ON) | NE | NE | 0.35 [-0.37, 1.06] | 0.68 [-0.06, 1.40] | Perseverative errors |
| Gotham et al. (1988) | 15 | 16 | 64.4 | 25.0 | -0.8 | -31.3 | 118.8 | - | - | - | - | M (B) | NE | NE | 0.93 [0.18, 1.67] | 0.86 [0.11, 1.59] | Perseverative errors |
| Graham et al. (2000) | 21 | 13 | 61.4 | 48.0 | -2.6 | 2.0 | 133.2 | - | - | 28.1 | -1.3 | ON | NE | NE | | | |
| Hanby et al. (2014) | 61 | 19 | 67.3 | 24.6 | 1.5 | -28.0 | 101.5 | 2.4 | 30.0 | - | - | ON | E | NE | | 0.77 [0.24, 1.29] | Perseverative errors |
| Hawkins et al. (2012) | 72 | 24 | 63.8 | 35.0 | -0.7 | -23.0 | 50.4 | 1.8 | 18.7 | - | - | ON | E | NE | | 0.71 [0.23, 1.18] | Perseverations |
| Hocherman et al. (2004) | 19 | 21 | 64.2 | 26.3 | 6.5 | -26.1 | 36.0 | 1.5 | - | - | - | M (ON) | E | NE | 0.55 [-0.08, 1.18] | 0.75 [0.10, 1.39] | Perseverative errors |
| Hozumi et al. (2000) | 15 | 13 | 65.4 | 53.3 | -0.8 | -8.2 | 67.2 | 2.1 | - | 27.9 | -0.3 | ON | NE | NE | 1.98 [1.05, 2.88] | 1.99 [1.06, 2.90] | Perseverative errors |
| Iijima et al. (2000) | 20 | 25 | 63.1 | 55.0 | -2.6 | 7.7 | 58.8 | 2.2 | - | - | - | ON | E | NE | 0 [-0.59, 0.59] | 0.07 [-0.52, 0.65] | Perseverative errors |
| Inzelberg et al. (2001) | 8 | 6 | 74.0 | 37.5 | 1.0 | -12.5 | 76.5 | 2.5 | - | - | - | OFF | E | NE | 1.48 [0.25, 2.67] | 1.21 [-0.02, 2.38] | Perseverative errors (%) |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ito & Kitagawa (2006) | 13 | 8 | 62.9 | 52.9 | -3.0 | -0.4 | 73.2 | 2.1 | - | 28.6 | -0.5 | ON | E | NE | 1.75 [0.69, 2.77] | 1.91 [0.83, 2.96] | Perseverative errors |
| Jahanshahi et al. (2002) | 13 | 12 | 57.0 | 23.1 | 1.9 | -51.9 | 174.0 | 2.9 | 42.1 | - | - | OFF | NE | NE | 0.77 [-0.05, 1.58] | 0.62 [-0.19, 1.42] | Perseverative errors |
| Katai et al. (2003) | 20 | 20 | 64.6 | 65.0 | 1.5 | 0.0 | 66.0 | 2.2 | 27.3 | 28.0 | -0.8 | M (ON) | E | NE | 0.75 [0.10, 1.38] | 0.29 [-0.33, 0.91] | Perseverative errors |
| Katsarou et al. (2004) | 45 | 40 | 59.3 | 31.1 | - | 3.6 | 73.2 | 2.5 | - | - | - | ON | E | NE | 0.35 [-0.08, 0.78] | 0.44 [0.01, 0.87] | Perseverative errors |
| Kaufman et al. (2016) | 14 | 12 | 63.3 | 21.0 | 1.6 | -21.0 | 121.2 | 2.4 | 25.4 | 29.2 | 0.5 | ON | E | NE | 1.37 [0.49, 2.21] | 1.15 [0.30, 1.97] | Perseverative responses |
| Krishna et al. (2014) | 76 | 43 | 66.3 | 30.3 | -0.6 | -2.3 | 102.4 | 2.6 | 21.4 | 27.6 | -0.7 | ON | E | NE | | | |
| Labudda et al. (2010) | 10 | 12 | 57.6 | 20.0 | -4.7 | -30.0 | 84.8 | 3.0 | - | - | - | ON | E | NE | 0.35 [-0.50, 1.19] | -0.05 [-0.89, 0.79] | Perseverations |
| Lange et al. (2016) | 32 | 35 | 62.6 | 34.4 | 0.4 | -22.7 | 93.6 | 2.0 | 19.7 | - | - | M (B) | E | NE | 0.68 [0.18, 1.17] | 0.84 [0.34, 1.34] | Perseverative errors |
| Lees & Smith (1983) | 30 | 30 | 58.9 | 36.7 | 5.0 | -6.6 | 28.8 | 1.8 | - | - | - | OFF (DN) | NE | NE | 0.55 [0.03, 1.06] | 0.78 [0.25, 1.30] | Perseverative errors |
| Leroi et al. (2012) | 102 | 33 | 63.1 | 28.7 | - | - | 95.2 | 2.2 | 28.1 | - | - | ON | E | NE | | | |
| Leroi et al. (2013) | 90 | 20 | 61.1 | 27.5 | 3.2 | -17.5 | 97.1 | 2.3 | 28.1 | - | - | ON | NE | NE | | 0.39 [-0.10, 0.87] | Perseverative errors |
| Levin et al. (1989) | 41 | 41 | 63.4 | 34.1 | - | - | 22.8 | 1.9 | - | - | - | M (ON) | E | NE | 0.37 [-0.07, 0.81] | 0.55 [0.11, 0.99] | Perseverative responses |
| Liozidou et al. (2012) | 73 | 48 | 61.2 | 38.4 | 1.6 | -5.4 | 124.8 | 2.0 | - | - | - | ON | E | NE | 2.13 [1.68, 2.59] | 1.53 [1.12, 1.94] | Perseverative errors |
| Lohmann et al. (2009) | 40* | 8 | 47.8 | 29.5 | -0.1 | -26.1 | 180.0 | 1.4 | 11.4 | 28.4 | -1.1 | ON | E | NE | 1.58 [0.75, 2.40] | 0.53 [-0.24, 1.30] | Perseverations |
| Marklund et al. (2009) | 18 | 10 | 65.1 | 50.0 | -4.0 | -10.0 | - | - | - | - | - | OFF (DN) | E | NE | 0.56 [-0.24, 1.34] | 0.50 [-0.29, 1.28] | Perseverative errors |
| McDowd et al. (2011) | 29 | 30 | 71.9 | - | -0.1 | - | - | 2.2 | 21.5 | 27.9 | -0.6 | - | E | NE | | 0.67 [0.14, 1.19] | Perseverative errors |
| Mignard et al. (2001) | 22 | 22 | 63.0 | 27.3 | -5.0 | 0.0 | 108.0 | 2.4 | 16.8 | - | - | ON | E | NE | 0.59 [-0.02, 1.19] | | |
| Mimura et al. (2006) | 18 | 20 | 68.9 | 72.2 | - | -7.8 | - | 2.5 | - | 27.8 | -1.1 | ON | E | NE | 0.96 [0.28, 1.63] | 0.49 [-0.16, 1.13] | Perseverative errors |
| Mioni et al. (2016) | 25 | 17 | 70.7 | 56.0 | 2.4 | 8.9 | - | - | 13.1 | 28.5 | -0.1 | - | E | E | 0.79 [0.14, 1.42] | 0.74 [0.10, 1.38] | Perseverative errors |
| Mohr et al. (1990) | 10 | 10 | 53.0 | 20.0 | 0.0 | 0.0 | 96.0 | 2.9 | - | - | - | M (ON) | E | NE | 0.27 [-0.61, 1.15] | 0.32 [-0.57, 1.19] | Perseverative responses |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mollion et al. (2003) | 18 | 9 | 57.6 | 33.3 | - | - | 96.0 | 2.0 | 16.0 | 29.3 | 0.1 | ON | NE | NE | 1.06 [0.20, 1.90] | 1.08 [0.21, 1.92] | Perseverative errors (%) |
| Moro dos Santos et al. (2010) | 21 | 22 | 74.0 | 57.1 | 5.0 | -24.0 | 84.0 | 1.5 | - | 25.8 | -2.2 | ON | E | NE | 0.05 [-0.55, 0.65] | -0.26 [-0.85, 0.35] | Perseverative errors |
| Müller et al. (2000) | 20 | 20 | 55.3 | 65.0 | -0.4 | 0.0 | 42.5 | 2.3 | - | - | - | M (B) | E | NE | 0.63 [-0.01, 1.26] | 1.10 [0.42, 1.76] | Perseverations |
| Muñiz Casado & Osuna Benavides (2007) | 18 | 18 | 71.7 | - | 1.8 | - | - | 1.8 | - | - | - | ON | E | E | 1.48 [0.73, 2.21] | 1.66 [0.89, 2.41] | Perseverative errors |
| Münte et al. (2015) | 12 | 12 | 66.5 | 58.3 | 0.8 | 0.0 | 124.8 | - | 22.3 | - | - | ON | E | NE | 0.27 [-0.53, 1.08] | | |
| Muslimovic et al. (2007) | 95 | 44 | 64.9 | 38.9 | 0.8 | -8.8 | 37.2 | 1.9 | 18.2 | 27.9 | -0.5 | M (ON) | E | NE | 0.53 [0.17, 0.89] | 0.63 [0.27, 1.00] | Perseverations |
| Nichelli et al. (1994) | 18 | 14 | 58.6 | - | -3.5 | - | - | 3.0 | - | - | - | M (ON) | NE | NE | 0.44 [-0.27, 1.14] | | |
| Nojszewska et al. (2009) | 46 | 14 | 65.6 | 37.0 | - | - | 93.6 | 2.5 | - | 26.7 | - | ON | E | NE | | 0.58 [-0.03, 1.19] | Perseverative errors (n) |
| Osternack Pinto (2005) - High H&Y | 17 | 18 | 64.0 | 70.0 | -1.4 | 60.0 | 178.8 | 3.2 | - | - | - | ON | NE | NE | 1.30 [0.56, 2.02] | 0.42 [-0.26, 1.08] | Perseverative responses |
| Osternack Pinto (2005) - Low H&Y | 19 | 18 | 67.4 | 60.0 | 2.0 | 50.0 | 109.2 | 2.3 | - | - | - | ON | NE | NE | 0.74 [0.07, 1.41] | 0.65 [-0.02, 1.31] | Perseverative responses |
| Paolo et al. (1995) | 181 | 187 | 68.9 | 34.8 | -0.8 | -28.3 | 67.1 | - | - | - | - | - | NE | NE | 0.86 [0.65, 1.08] | 0.65 [0.44, 0.86] | Perseverative errors |
| Passamonti et al. (2013) | 16 | 13 | 59.6 | 25.0 | -0.6 | -16.7 | 36.7 | 1.8 | 21.5 | - | - | OFF | E | NE | | 0.12 [-0.61, 0.85] | Perseverative errors |
| Pavlova et al. (2014) | 46 | 20 | 69.6 | 30.4 | 0.1 | 0.4 | 70.7 | - | 31.9 | - | - | M (U) | NE | NE | 2.65 [1.95, 3.34] | | |
| Pell & Leonard (2003) | 21 | 21 | 61.7 | 47.6 | -0.2 | 0.0 | 46.8 | 2.0 | 14.5 | - | - | M (ON) | E | NE | 0.38 [-0.23, 0.99] | 0.46 [-0.15, 1.07] | Perseverative errors (%) |
| Pellicano et al. (2012) | 13 | 13 | 58.8 | 31.0 | -1.5 | -7.0 | 51.6 | 1.9 | 18.5 | 28.4 | -1.1 | ON | E | NE | | -0.12 [-0.89, 0.65] | Perseverative errors |
| Pellicano et al. (2015) | 84 | 84 | 63.3 | 38.1 | 0.2 | 0.0 | 12.0 | 1.6 | 15.1 | 28.6 | -0.6 | OFF (DN) | E | NE | 0.36 [0.06, 0.67] | 0.45 [0.14, 0.76] | Perseverative errors |
| Perfetti et al. (2010) | 25 | 24 | 69.8 | 48.0 | -3.1 | -22.8 | - | 2.2 | 19.9 | 27.0 | - | ON | E | NE | | 0.94 [0.34, 1.53] | Perseverative responses |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Péron et al. (2010a) | 44 | 30 | 61.4 | 43.2 | 2.4 | -6.8 | 139.2 | 1.3 | 11.2 | - | - | ON | NE | NE | 0.94 [0.12, 1.75] | 0.69 [-0.11, 1.47] | Perseverative errors (n) |
| Péron et al. (2010b) | 21 | 21 | 59.5 | 52.4 | 1.3 | 0.0 | 132.0 | 1.3 | 9.5 | - | - | ON | NE | NE | 0.64 [0.16, 1.11] | 0.93 [0.44, 1.42] | Perseverations |
| Péron et al. (2010c) | 13 | 13 | 53.3 | 38.5 | - | 0.0 | 126.0 | 1.2 | 8.8 | - | - | ON | E | NE | 0.69 [0.06, 1.31] | 0.50 [-0.12, 1.11] | Perseverations |
| Péron et al. (2014) - Advanced PD | 15 | 15 | 59.5 | 66.7 | 3.6 | 0.0 | 133.2 | 1.3 | - | - | - | ON | NE | NE | 0.28 [-0.44, 1.00] | 0.49 [-0.24, 1.22] | Perseverative errors (n) |
| Péron et al. (2014) - Early PD | 15 | 15 | 60.3 | 66.7 | 4.4 | 0.0 | 33.6 | 0.6 | - | - | - | ON | NE | NE | 0.37 [-0.35, 1.09] | 0.50 [-0.23, 1.22] | Perseverative errors (n) |
| Perretta et al. (2005) - High H&Y | 16 | 17 | 77.7 | 50.0 | 5.1 | 7.9 | - | 3.3 | 27.2 | 28.1 | -0.8 | ON | E | NE | 1.17 [0.42, 1.90] | 0.34 [-0.35, 1.02] | Perseverative errors (%) |
| Perretta et al. (2005) - Low H&Y | 14 | 17 | 72.4 | 43.8 | -0.2 | 1.7 | - | 2.1 | 11.3 | 29.0 | 0.1 | ON | E | NE | 0.60 [-0.13, 1.32] | 0.10 [-0.61, 0.80] | Perseverative errors (%) |
| Petrova et al. (2010) | 23 | 25 | 67.9 | - | 0.5 | - | 75.6 | 2.3 | 21.2 | 28.2 | -0.4 | - | E | NE | 0.84 [0.24, 1.42] | 0.88 [0.28, 1.47] | Perseverations |
| Petrova et al. (2012) - Mild dementia | 22 | 26 | 73.1 | - | 4.4 | - | 103.2 | 2.7 | 29.3 | 21.9 | -6.6 | - | NE | NE | 0.98 [0.37, 1.57] | | |
| Petrova et al. (2012) - Very mild dementia | 36 | 26 | 69.8 | - | 1.1 | - | 98.4 | 2.6 | 30.5 | 26.4 | -2.1 | - | E | NE | 0.62 [0.10, 1.13] | | |
| Pillon et al. (1996) | 20 | 14 | 62.4 | - | -1.9 | - | 97.2 | 2.5 | 18.1 | 28.9 | -0.3 | ON | E | E | 0.56 [-0.14, 1.25] | 1.04 [0.30, 1.76] | Perseverations |
| Pirogovsky-Turk et al. (2017) | 68 | 30 | 67.0 | 35.3 | -2.1 | -21.4 | 73.2 | 2.0 | 23.9 | - | - | ON | E | NE | | 0.07 [-0.36, 0.50] | Perseverative responses |
| Poletti et al. (2012) | 126 | 100 | 66.6 | 37.3 | -0.2 | -15.7 | 166.8 | - | 16.9 | 27.5 | -0.2 | OFF (DN) | E | NE | 0.15 [-0.11, 0.42] | 0.09 [-0.17, 0.36] | Perseverative errors |
| Possin (2007) - Experiment1 | 18 | 15 | 67.4 | 44.4 | 0.1 | 0.0 | 73.2 | 2.0 | 21.5 | - | - | ON | E | NE | 0.17 [-0.52, 0.85] | -0.01 [-0.25, 0.25] | Perseverative responses |
| Possin (2007) - Experiment2 | 17 | 12 | 67.0 | 38.9 | -2.4 | -11.1 | 64.8 | 2.1 | 23.9 | - | - | ON | E | NE | -0.13 [-0.93, 0.67] | -0.25 [-0.99, 0.49] | Perseverative responses |
| Possin (2007) - Experiment3 | 15 | 10 | 69.5 | 40.0 | 2.8 | -6.7 | 73.2 | 2.2 | 21.7 | - | - | ON | E | NE | 0.41 [-0.40, 1.22] | 0.61 [-0.21, 1.43] | Perseverative responses |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pozzi et al. (1994) - Demented | 13 | 10 | 70.4 | 31.0 | 1.1 | -19.0 | 57.6 | - | - | 19.6 | -8.5 | ON | NE | NE | | |
| Pozzi et al. (1994) – Non-demented | 34 | 10 | 63.5 | 62.0 | -5.8 | 12.0 | 64.8 | - | - | 27.5 | -0.6 | ON | E | NE | | |
| Price (2005) | 17 | 18 | 66.8 | 59.0 | -1.4 | -3.0 | 98.4 | 2.3 | - | 28.0 | -0.9 | ON | E | NE | | 1.25 [0.51, 1.97] | Perseverative errors |
| Price (2006) | 16 | 17 | 66.4 | 62.5 | 0.0 | 3.7 | 98.4 | - | - | - | - | ON | E | NE | | 1.02 [0.28, 1.74] | Perseverative errors |
| Price (2010) | 15 | 12 | 67.7 | 33.3 | 3.5 | 0.0 | 77.6 | 1.9 | - | 28.4 | 0.5 | ON | E | E | | 0.28 [-0.49, 1.04] | Perseverative errors |
| Price & Shin (2009) | 22 | 10 | 71.7 | 36.1 | 1.2 | -23.9 | 79.1 | 1.8 | 12.6 | 28.6 | -0.6 | ON | E | NE | 0.23 [-0.52, 0.98] | 0.71 [-0.06, 1.47] | Perseverative errors |
| Puertas-Martin et al. (2016) | 32 | 32 | 67.7 | 40.6 | -0.2 | -3.1 | 76.8 | 2.5 | 14.9 | - | - | ON | E | NE | | 0.12 [-0.37, 0.61] | Perseverations |
| Ravizza & Ciranni (2002) | 9 | 13 | 68.0 | - | 0.0 | - | 153.6 | 2.6 | - | - | - | ON | NE | NE | 0.94 [0.03, 1.83] | | |
| Roca et al. (2012) | 32 | 22 | 62.3 | - | 3.0 | - | 17.6 | 1.5 | - | - | - | M (B) | E | NE | 0.77 [0.21, 1.33] | | |
| Rosen et al. (2013) | 19 | 20 | 65.2 | 63.2 | 3.1 | -1.8 | 69.5 | 2.5 | - | 28.6 | -1.0 | ON | E | NE | | | |
| Rosen et al. (2015) | 20 | 23 | 67.5 | 30.0 | -0.8 | -4.8 | 100.8 | 2.5 | - | 28.8 | -0.3 | ON | E | NE | | 1.68 [0.94, 2.41] | Perseverations |
| Rouillard et al. (2017) | 49 | 47 | 66.3 | 44.9 | 2.4 | -6.2 | 76.0 | 1.6 | - | 27.8 | -1.1 | M (ON) | E | NE | | 0.23 [-0.17, 0.63] | Perseverative errors |
| Sagar et al. (1991) | 56 | 32 | 60.1 | 48.2 | 1.6 | -1.8 | 13.2 | - | - | - | - | OFF (DN) | NE | NE | 0.34 [-0.09, 0.78] | -0.20 [-0.63, 0.24] | Perseverations |
| Sánchez et al. (2002) | 33 | 46 | 69.7 | 48.5 | -0.8 | -10.2 | 28.7 | 2.0 | - | - | - | ON | NE | NE | 1.38 [0.88, 1.88] | | |
| Schmidt (2013) | 62 | 32 | 64.5 | 38.7 | 1.6 | -20.7 | 70.8 | 2.1 | 18.0 | 28.7 | -0.1 | - | E | NE | 0.76 [0.32, 1.20] | 0.63 [0.20, 1.07] | Perseverations |
| Smith & McDowall (2006a) | 31 | 28 | 63.0 | 29.0 | -2.7 | -13.9 | 81.5 | 2.3 | - | 28.8 | -0.2 | M (ON) | E | NE | 0.29 [-0.22, 0.80] | 0.44 [-0.08, 0.96] | Perseverative errors |
| Smith & McDowall (2006b) | 18 | 22 | 58.1 | 33.3 | 0.1 | -12.2 | 81.8 | 2.3 | - | 28.9 | -0.1 | ON | E | NE | -0.02 [-0.53, 0.49] | -0.37 [-1.00, 0.26] | Perseverative errors |
| Smith & McDowall (2011) | 16 | 18 | 62.7 | 26.7 | 1.7 | -12.2 | 63.7 | 2.2 | - | 28.8 | 0.0 | ON | E | NE | -0.01 [-0.54, 0.53] | -0.52 [-1.20, 0.17] | Perseverative errors |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Stamenovic et al. (2003) | 30 | 15 | 59.2 | 33.3 | - | - | 180.0 | 1.3 | - | 27.8 | -1.0 | OFF | E | NE | 1.97 [1.21, 2.70] | | |
| Stefanova et al. (2001) | 39 | 31 | 49.3 | 38.5 | 1.0 | -26.0 | 57.6 | 1.6 | - | - | - | ON | E | E | 0.68 [0.19, 1.17] | 0.95 [0.45, 1.44] | Perseverative errors |
| Strohmaier (2016) | 55 | 19 | 66.1 | 49.1 | 1.5 | 1.7 | 103.0 | - | 20.6 | 27.9 | -1.0 | - | NE | NE | 1.34 [0.77, 1.91] | 1.00 [0.45, 1.54] | Perseverations |
| Taylor et al. (1986) | 40 | 40 | 60.5 | 37.5 | -0.2 | -10.0 | 79.4 | 2.3 | - | - | - | M (ON) | E | NE | 0.72 [0.27, 1.17] | | |
| Tomer et al. (2002) | 28 | 19 | 66.4 | 35.7 | -0.7 | -11.7 | 0.0 | - | 17.1 | 28.5 | -0.5 | OFF (DN) | NE | NE | 1.04 [0.41, 1.66] | 0.76 [0.16, 1.36] | Perseverative errors |
| Torralva et al. (2015) | 32 | 22 | 62.3 | - | 3.0 | - | - | 1.5 | - | - | - | - | NE | NE | | | |
| Tröster et al. (1995) | 83 | 43 | 69.1 | 35.2 | -0.1 | -3.4 | 66.0 | 2.1 | 20.6 | - | - | ON | NE | NE | 0.99 [0.60, 1.38] | 0.82 [0.44, 1.20] | Perseverations |
| Tröster et al. (2006) | 61 | 144 | 68.6 | 29.2 | -2.9 | -21.8 | 71.2 | 2.3 | - | - | - | ON | NE | NE | 0.64 [0.33, 0.94] | | |
| Vance (1990) | 19 | 19 | 67.3 | 31.6 | -2.0 | -31.6 | - | 2.0 | - | 28.7 | -0.5 | ON | E | NE | 1.05 [0.36, 1.72] | 0.86 [0.19, 1.52] | Perseverations |
| Venneri et al. (1997) | 25 | 22 | 60.4 | - | -1.9 | - | 36.6 | 2.0 | - | 28.7 | -0.7 | ON | E | NE | 0.86 [0.26, 1.46] | 0.85 [0.25, 1.44] | Perseverative errors (n) |
| Vincente et al. (2011) - Advanced PD | 18 | 15 | 60.3 | 55.6 | 3.0 | 2.3 | 138.6 | 1.4 | - | - | - | ON | E | NE | 0.68 [-0.03, 1.39] | 0.16 [-0.53, 0.84] | Perseverative errors (n) |
| Vincente et al. (2011) - Early PD | 15 | 15 | 62.3 | 66.7 | 5.1 | 13.4 | 29.8 | 0.8 | - | - | - | ON | E | NE | 0.35 [-0.37, 1.07] | 0.28 [-0.44, 1.00] | Perseverative errors (n) |
| Werheid et al. (2007) | 14 | 16 | 62.5 | 42.9 | 0.1 | -0.9 | 67.6 | - | - | - | - | ON | E | E | 0.53 [-0.20, 1.26] | | |
| Wild et al. (2012) | 18 | 18 | 69.3 | 55.6 | -0.1 | 0.0 | 100.7 | 2.0 | 16.2 | 26.4 | -0.7 | ON | E | NE | 0.69 [0.01, 1.35] | 0.32 [-0.34, 0.98] | Perseverative errors |
| Willemsen et al. (2008) | 20 | 20 | 64.5 | 40.0 | 0.2 | 0.0 | 38.4 | - | 10.8 | - | - | ON | NE | NE | | | |
| Willemsen et al. (2009) | 14 | 14 | 58.9 | 50.0 | -0.1 | - | 0.0 | - | 12.5 | - | - | OFF (DN) | NE | NE | 0.55 [-0.21, 1.30] | 0.43 [-0.33, 1.17] | Perseverative errors |
| Witt et al. (2002) | 23 | 20 | 60.4 | 52.2 | 0.5 | 17.2 | - | 2.3 | 17.8 | 28.2 | -0.8 | ON | E | NE | 1.16 [0.51, 1.81] | | |
| Witt et al. (2006a) | 22 | 22 | 58.0 | 27.3 | 1.1 | -13.6 | 97.1 | - | 16.6 | - | - | ON | E | NE | 0.77 [0.16, 1.38] | 0.40 [-0.20, 1.00] | Perseverative errors |
| Witt et al. (2006b) | 20 | 20 | 59.3 | 30.0 | 0.3 | -10.0 | 39.0 | 2.0 | 15.4 | - | - | ON | E | NE | 1.11 [0.43, 1.77] | 0.63 [-0.01, 1.26] | Perseverative errors |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Woods & Tröster (2003) | 36 | 18 | 69.5 | 33.0 | 0.8 | 0.0 | 71.0 | 2.1 | - | - | - | ON | E | NE | 0.48 [-0.10, 1.05] | 0.92 [0.33, 1.51] | Perseverative errors |
| Yu et al., (2010) | 55 | 30 | 62.5 | 32.7 | -1.7 | -20.6 | 43.4 | 1.5 | 16.7 | 28.3 | -0.1 | - | E | NE | 0.54 [0.09, 0.99] | 0.20 [-0.25, 0.64] | Perseverative errors |
| Yu et al. (2012a) | 94 | 84 | 61.6 | 37.2 | 0.6 | -6.8 | 48.4 | 1.5 | - | - | - | - | NE | NE | 0.76 [0.46, 1.07] | 0.47 [0.17, 0.76] | Perseverative errors |
| Yu et al. (2012b) | 39 | 40 | 62.7 | 35.9 | 0.8 | -14.1 | 51.6 | 1.6 | 18.9 | 27.9 | 0.0 | ON | E | NE | 0.30 [-0.15, 0.74] | 0.06 [-0.38, 0.50] | Perseverative errors |
| Zeng et al. (2002) | 18 | 16 | 63.9 | 33.3 | 2.7 | -10.5 | 54.7 | 1.7 | - | 28.7 | -0.2 | ON | E | NE | 1.37 [0.61, 2.12] | 1.46 [0.69, 2.21] | Perseverative errors |
| | $N_{PD}$ | $N_{HC}$ | age | %F | $\Delta_{age}$ | $\Delta_{\%F}$ | dur | HY | UPDRS | MMSE | $\Delta_{MMSE}$ | | | | | | |
| Mean | 30 | 24 | 64.0 | 41.8 | 0.5 | -7.0 | 76.0 | 2.1 | 20.4 | 28.0 | -0.9 | | | | | | |
| Standard deviation | 24 | 20 | 5.0 | 12.9 | 2.2 | 14.3 | 39.0 | 0.5 | 6.8 | 1.4 | 1.3 | | | | | | |
| % reported | 100 | 100 | 100 | 84 | 96 | 81 | 88 | 79 | 50 | 50 | 48 | 93 | 100 | 100 | 80 | 79 | 79 |

*Note*. The column "Selected measure of perseveration" displays the description of the selected perseveration measures as used by the authors of the original paper. age = mean age of participants in the patient group in years, %F = proportion of female participants in the patient group, Δ = Difference patients - controls, dur = disease duration in the patient group, HY = Hoehn & Yahr-stage in the patient group, UPDRS = mean score on the motor scale of the Unified Parkinson's Disease Rating Scale in the patient group, MMSE = mean score of the Mini-Mental State Examination in the patient group, med = medication status in the patient group (OFF (DN) = OFF (de novo), OFF (DW) = OFF (dopamine withdrawal), OFF (M) = OFF (mixed = de novo and withdrawal), M (ON) = Mixed (Majority ON), M (OFF) = Mixed (Majority OFF), M (U) = Mixed (Unknown), M (B) = Mixed (Balanced), dem = dementia status of the patient group: E = Excluded, NE = Not Excluded, dep = depression status of the patient group: E = Excluded, NE = Not Excluded "-" = data not available, *sample size differs across different WCST measures (categories: $n$ = 40, perseverations: $n$ = 39)

Table 2.
*Results of the meta-analyses comparing WCST performance between patients with Parkinson's disease and healthy control participants*

| | Categories | Perseverations | Perseverative errors (n) | Perseverative errors (%) | Perseverative responses | Non-perseverative errors | Total Errors | Trials to criterion | Failures to maintain set | Conceptual Level responses | Global score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Number of samples (*k*)** | 144 | 143 | 85 | 22 | 29 | 31 | 60 | 18 | 27 | 13 | 13 |
| **Significant effects (%)** | 59.03 | 46.85 | 45.88 | 50.00 | 48.28 | 58.06 | 63.33 | 22.22 | 25.93 | 53.85 | 61.54 |
| **Total $N_{PD}$** | 4166 | 4324 | 2651 | 513 | 995 | 1261 | 1668 | 714 | 786 | 634 | 449 |
| **Total $N_{HC}$** | 3561 | 3417 | 2146 | 430 | 800 | 956 | 1399 | 594 | 704 | 475 | 239 |
| **Average effect size Cohen's *d*** | 0.74 | 0.57 | 0.56 | 0.60 | 0.59 | 0.58 | 0.78 | 0.38 | 0.29 | 0.68 | 0.77 |
| **[95% CI]** | [0.67, 0.82] | [0.50, 0.64] | [0.46, 0.65] | [0.41, 0.79] | [0.47, 0.71] | [0.48, 0.67] | [0.65, 0.91] | [0.25, 0.51] | [0.13, 0.45] | [0.37, 0.99] | [0.54, 1.01] |
| ***Q*** | 331.55* | 290.67* | 182.03* | 38.87* | 39.45 | 33.00 | 159.70* | 20.57 | 52.07* | 59.94* | 21.56* |
| ***I²*** | 56.27 | 51.15 | 53.85 | 45.97 | 23.94 | 3.04 | 63.06 | 7.62 | 50.07 | 76.64 | 44.34 |

*Note*. *$p$ = .05.

Table 3.
*Assessment of the potential impact of publication bias in our meta-analysis on PD-related WCST performance deficits*

| | **Categories** | **Perseverations** | **Perseverative errors (n)** | **Perseverative errors (%)** | **Perseverative responses** | **Non-perseverative errors** | **Total Errors** | **Trials to criterion** | **Failures to maintain set** | **Conceptual Level responses** | **Global score** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $d_{\text{random-effects meta-analysis}}$ | 0.74 | 0.57 | 0.56 | 0.60 | 0.59 | 0.58 | 0.78 | 0.38 | 0.29 | 0.68 | 0.77 |
| **[95% CI]** random-effects meta-analysis | [0.67, 0.82] | [0.50, 0.64] | [0.46, 0.65] | [0.41, 0.79] | [0.47, 0.71] | [0.48, 0.67] | [0.65, 0.91] | [0.25, 0.51] | [0.13, 0.45] | [0.37, 0.99] | [0.54, 1.01] |
| $d_{\text{non-significant studies}}$ | 0.34 | 0.25 | 0.25 | 0.46 | 0.32 | 0.38 | 0.36 | 0.25 | 0.11 | 0.24 | 0.71 |
| **[95% CI]**$_{\text{non-significant studies}}$ | [0.26, 0.42] | [0.18, 0.38] | [0.16, 0.33] | [0.20, 0.71] | [0.17, 0.47] | [0.21, 0.54] | [0.23, 0.49] | [0.12, 0.38] | [0.00, 0.23] | [0.02, 0.50] | [0.39, 1.03 |
| $\tau_{\text{Begg \& Mazumar}}$ | .20* | .13* | .17* | .23 | .16 | .32* | .22* | .23 | .08 | .23 | .36 |
| $d_{\text{regression}}$ | 0.69 | 0.49 | 0.46 | 0.46 | 0.52 | 0.40 | 0.73 | 0.14 | 0.19 | 0.80 | 0.35 |
| **[95% CI]**$_{\text{regression}}$ | [0.54, 0.84] | [0.36, 0.63] | [0.32, 0.65] | [0.04, 0.95] | [0.30, 0.74] | [0.25, 0.56] | [0.42, 1.05] | [-0.07, 0.35] | [-0.10, 0.49] | [0.25, 1.35] | [-0.09, 0.81] |
| $d_{\text{moderate one-tailed selection}}$ | 0.68 | 0.50 | 0.49 | 0.53 | 0.55 | 0.55 | 0.71 | 0.34 | 0.21 | 0.58 | 0.69 |
| $d_{\text{severe one-tailed selection}}$ | 0.61 | 0.39 | 0.37 | 0.43 | 0.51 | 0.52 | 0.62 | 0.28 | -1.07 | -1.24 | 0.67 |
| $d_{\text{moderate two-tailed selection}}$ | 0.70 | 0.53 | 0.52 | 0.56 | 0.56 | 0.55 | 0.74 | 0.35 | 0.27 | 0.63 | 0.70 |
| $d_{\text{severe two-tailed selection}}$ | 0.64 | 0.47 | 0.47 | 0.49 | 0.52 | 0.52 | 0.70 | 0.30 | 0.24 | 0.56 | 0.69 |

*Note*. The first two rows present the results from our random-effects meta-analysis for comparison. The following two rows display the effect sizes and their confidence intervals (CIs) for those studies that reported non-significant results. Begg and Mazumar's rank correlation coefficient ($\tau_{Begg\ \&\ Mazumar}$) describes the association between effect sizes and their standard errors across all included samples. $d_{regression}$ is the intersect of the weighted linear regression model predicting effect sizes from the inverse of sample sizes. The final four rows display the results of the four selection bias models proposed by Vevea and Woods (2005).

Table 4.
Results of the meta-regression analyses conducted to examine the role of potential moderators of PD-related WCST performance deficits

| | Categories | | | | | Perseverations | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Continuous moderators | β | 95% CI | df | *t* | *p* | β | 95% CI | df | *t* | *p* |
| Age | .02 | [-.07, .10] | 141 | 0.38 | .708 | -.01 | [-.08, .07] | 140 | -0.19 | .846 |
| Percent female patients | -.03 | [-.12, .06] | 120 | -0.65 | .514 | .01 | [-.07, .09] | 124 | 0.23 | .816 |
| Disease duration | .09* | [.01, .17] | 124 | 2.13 | .036 | .07 | [-.01, .14] | 126 | 1.68 | .092 |
| Hoehn & Yahr | .07 | [-.02, .16] | 112 | 1.58 | .118 | .05 | [-.04, .13] | 114 | 1.10 | .275 |
| UPDRS motor score | .16* | [.06, .27] | 69 | 3.08 | .003 | .03 | [-.06, .12] | 71 | 0.62 | .537 |
| MMSE | .02 | [-.07, .11] | 69 | 0.46 | .646 | -.04 | [-.14, .06] | 71 | -0.81 | .420 |
| PD-HC difference age | .06 | [-.02, .14] | 135 | 1.46 | .146 | .04 | [-.04, .11] | 134 | 0.89 | .377 |
| PD-HC difference percent female participants | -.02 | [-.11, .07] | 115 | -0.39 | .699 | -.03 | [-.10, .05] | 119 | -0.69 | .493 |
| PD-HC difference MMSE | -.05 | [-.14, .04] | 68 | -1.10 | .274 | -.06 | [-.18, .07] | 68 | -0.93 | .354 |
| Categorical moderators | $\chi^2$ | | df | | *p* | $\chi^2$ | | df | | *p* |
| Dementia status | 3.28 | | 1 | | .070 | 1.29 | | 1 | | .255 |
| Depression status | 0.05 | | 1 | | .820 | 2.69 | | 1 | | .101 |
| Medication status | 8.24* | | 2 | | .016 | 4.74 | | 2 | | .094 |
| Publication status | 0.15 | | 1 | | .701 | 0.39 | | 1 | | .532 |

*Note*: *$p < .05$.