Arithmetic deficits in Parkinson's Disease?

– A Registered Report

Hannah D. Loenneker[1]*, Inga Liepelt-Scarfone[2,3], Klaus Willmes[4], Hans-Christoph Nuerk[1,5,6], & Christina Artemenko[1,5]

[1]Department of Psychology, University of Tuebingen, Tuebingen, Germany

[2]Hertie Institute for Clinical Brain Research and German Centre for Neurodegenerative Diseases, Tuebingen, Germany

[3]IB-Hochschule für Gesundheit und Soziales, Stuttgart, Germany

[4]Department of Neurology, RWTH Aachen University, University Hospital, Aachen, Germany

[5]LEAD Graduate School & Research Network, University of Tuebingen, Tuebingen, Germany

[6]German Center for Mental Health, Tuebingen, Germany

*Corresponding author:

E-Mail: hannah-dorothea.loenneker@uni-tuebingen.de (HDL)

The current article is a stage 2 registered report and is linked to the following stage 1 registered report: Loenneker, H. D., Liepelt-Scarfone, I., Willmes, K., Nuerk, H.-C., & Artemenko, C. (2022). Arithmetic Deficits in Parkinson's Disease? A Registered Report. Stage 1 In-Principle Acceptance by Peer Community In Registered Reports. https://osf.io/nb5fj

This stage 2 manuscript has been submitted to Peer Community In Registered Reports for stage 2 acceptance on 2025/10/14 and is currently under review.

Abstract

Elderly people and patients with neurodegenerative diseases such as Parkinson's Disease (PD) immensely rely on arithmetic skills to lead an independent life. Activities such as medication management, financial transactions or using public transport require intact abilities to manipulate numbers with different arithmetic operations. However, research on cognitive deficits in PD has been focussing on domain-general functions such as executive functions, attention or working memory so far – largely neglecting potential domain-specific aspects of numerical cognition (e.g., carry or problem size effect). These aspects should be addressed, as PD-immanent deterioration of domain-specific numerical areas and domain-general functions suggests mechanisms of both primary and secondary (mediated by other cognitive deficits) arithmetic deficits, respectively. The current study systematically investigated arithmetic performance and effects in PD patients differing in cognitive impairment for the first time, targeting domain-specific cognitive representations of arithmetic as well as the influence of domain-general factors. Besides healthy controls (HC), PD patients with normal cognition (PD-NC) and PD patients with mild cognitive impairment (PD-MCI) were compared in arithmetic performance in the four basic operations (addition, subtraction, multiplication, division). The deficits consisted of problems in two-digit addition and subtraction and PD-MCI, as well as delayed arithmetic fact retrieval for single-digit multiplication and division in PD-NC and PD-MCI. Deficits were not domain-specific, as arithmetic effects (problem size, carry, and borrow effects) did not differ between the groups. Discriminant analyses showed that performance in addition, multiplication and division tasks could differentiate between PD-NC and PD-MCI. The study results help us to understand the underlying mechanisms of arithmetic deficits faced by PD patients in daily life.

Words: 260

Keywords: Parkinson's disease, mild cognitive impairment, arithmetic operation, calculation, place × value system

Introduction

In a society shaped by demographic change and consequences of ageing communities, supporting independence of the elderly is crucial. Daily activities of the elderly such as planning a trip with public transport, keeping track of the financial situation, and managing one's own medication all share a common prerequisite for completion: (intact) arithmetic skills (Arcara et al., 2019; Bangma et al., 2021; Delazer et al., 2013). Deficits in arithmetic skills can be acquired (i.e., acalculia) or result from developmental deficits (i.e., dyscalculia, Willmes et al., 2013). Despite the importance of arithmetic skills for the elderly, acalculia research is missing in this societal group (Ardila & Rosselli, 2002). What is even more important is the lack of research considering patients with neurodegenerative diseases, who show deficits in several arithmetic domains in clinical practice (e.g., Cappelletti et al., 2005; Delazer et al., 2019; Kalbe & Kessler, 2002), as for example patients with Alzheimer's Disease (AD, for reviews see Girelli & Delazer, 2001; Kalbe & Kessler, 2002). While at least some literature on arithmetic and associated financial decision-making exists for AD (Bangma et al., 2021), Parkinson's Disease (PD) has been neglected in systematic research on arithmetic deficits.

Therefore, the current study is part of a broader research project on numerical cognition in PD, which aims to assess basic number processing (Loenneker et al., 2024) as well as arithmetic skills (the current study), to obtain a comprehensive overview of how basic and advanced numerical functions look like in PD patients with different levels of global cognitive function. Both studies will be conducted with the same patients and within two joint experimental sessions.

Parkinson's Disease

With to date over 6 million globally affected patients, PD is the second most common neurodegenerative disease world-wide (Dorsey et al., 2018). Prevalence differs depending on sex and age with a higher incidence in men and the elderly (Hirsch et al., 2016). The increasing prevalence of PD and associated health costs and care demands will have a huge impact on the public health system in the future (Dorsey et al., 2018; Kowal et al., 2013). PD's cardinal motor symptoms are hypo- and bradykinesia alongside at least rigidity and/ or resting tremor (Postuma et al., 2015). In addition, the disease is characterised by the presence of a variety of non-motor symptoms such as dysautonomia (e.g., incontinence, sleep disorders), psychiatric (e.g., depression, Jankovic, 2008) or cognitive disorders such as mild cognitive impairment (PD-MCI, Litvan et al., 2012) and dementia (PDD, Aarsland et al., 2017). PD's cognitive symptoms are consensually defined as deficits in the domains of executive functions, working memory and attention, memory, language, and visuo-spatial functions (Aarsland et al., 2017; Litvan et al., 2012).

The continuum of cognitive deficits in PD can be further differentiated into normal cognition (PD-NC), the initial stage of clinically significant cognitive disorders (PD-MCI) and PDD (with additional impairment in activities of daily living = ADL, Emre et al., 2007). Especially cognitive impairment is associated with a more rapid disease progression, mortality and death (de Lau et al., 2014; Pigott et al., 2015). Previous studies in neurological patients have shown that general cognitive abilities as well as domain-specific numerical functions influence arithmetic functions (Ardila & Rosselli, 2002; Willmes et al., 2013). The overlap of brain areas underlying arithmetic (Klein et al., 2016) and degenerated areas in PD (Braak et al., 1996, 2006) emphasise the need for theory-driven investigations.

Arithmetic across aging

Arithmetic deficits in PD need to be differentiated from non-pathological aging effects, but lifespan psychological research mostly focuses on global cognitive decline (e.g., Cohen-Mansfield et al., 2018; Deary et al., 2013; Lindenberger & Baltes, 1997; Tucker-Drob, 2019). Neurocognitive models of number processing (i.e., the Triple Code Model and its extensions, Dehaene et al., 2003; Dehaene & Cohen, 1995a; Klein et al., 2016) lack hypotheses on ageing processes, but generally assume different representations of numerosity being associated with specific subcortical (e.g., basal ganglia, hippocampus) as well as cortical (e.g., fronto-parietal network) structures.

The Triple Code Model postulates 1) a *visual number form,* 2) a *semantic magnitude* and 3) a *verbal representation* as the core representations of numerosity (Dehaene et al., 2003; Dehaene & Cohen, 1995a). These number-related processes are organized in three parietal circuits, being the horizontal segment of the intraparietal sulcus (for notation-independent quantity processing), the left angular gyrus area with further left-hemispheric perisylvian areas (for verbal manipulations) and a bilateral posterior superior parietal system (for attention regarding spatial dimensions, Dehaene et al., 2003). As a fourth representation, simple multiplication and division tasks require *arithmetic fact retrieval* from verbal long-term memory facilitating automatized mental calculation (as hypothesized e.g., in the interacting neighbours model, Verguts & Fias, 2005). These have been complemented by *place × value integration* (identification, activation and computation of distinct places and corresponding values) for multi-digit numbers (Klein et al., 2016; Nuerk et al., 2015).

Representations postulated by these models can be operationalised in an effect-based (i.e., numerical representation is indexed by a numerical effect) as opposed to a task-based (i.e., numerical representation is indexed by a numerical task) approach (see Moeller et al., 2011 for further elaboration on this distinction). Regarding arithmetic, magnitude and verbal representations as well as place × value integration are crucial underlying mechanisms, with distinct accentuations for the four basic arithmetic operations and several related numerical effects. For instance, place × value computation is relevant for

carrying in addition (unit sum of the operands ≤ 10, e.g., 32 + 48) and borrowing in subtraction (unit of the subtrahend > unit of the minuend, e.g., 51 - 27), which lead to longer reaction times (RT) and lower accuracy (ACC; Artemenko et al., 2018). While carry and borrow effects define task complexity in addition and subtraction, task complexity in multiplication and division can be defined respectively based on problem size of the product or divisor, with the problem size effect showing faster RT and higher ACC for smaller (i.e., easier) than larger problems (Domahs et al., 2006; Zbrodoff & Logan, 2005). Additionally, erroneous place × value integration is for example revealed by operand-related errors (the erroneous answer belongs to the multiplication table of one of the operands, e.g., 7 × 8 = 48 instead of 56, result and error have the same digit at the same place-value position in common, e.g., 7 × 8 = 54) in multiplication and division (Domahs et al., 2006, 2007; Stazyk et al., 1982; Verguts & Fias, 2005). These effects can be explained with operands and solutions in multi-digit arithmetic being represented in network structures (Domahs et al., 2006; Verguts & Fias, 2005).

However, domain-specific numerical representations such as number magnitude representation, place-value integration or knowledge of mathematical procedures cannot explain arithmetic performance entirely. Domain-general functions such as attention, working memory, or executive functions are an additional prerequisite for successful arithmetic processing and deficits therein can give rise to arithmetic deficits comparable to secondary acalculia (termed secondary arithmetic deficits hereafter, Knops et al., 2017). For instance, carry and borrow effects are strongly related to working memory capacity (Imbo, Vandierendonck, & De Rammelaere, 2007; Imbo, Vandierendonck, & Vergauwe, 2007) and interference effects in multiplication indicate inhibition as a crucial prerequisite of arithmetic (Archambeau et al., 2019; De Visscher & Noël, 2014; Domahs et al., 2006; Verguts & Fias, 2005).

In the fields of developmental numerical cognition, the few studies describing arithmetic processes in healthy ageing showed differential results. While performance of elderly people does not seem to be declining significantly in arithmetic fact retrieval and magnitude representation (Archambeau et al., 2019; Kaufmann et al., 2008), they do show increased error rates and differences in neural activation in a numerical Stroop task (requiring inhibition; Wood et al., 2009), and less efficient strategy use in other arithmetic operations when compared to young adults (Hinault & Lemaire, 2016a; Uittenhove & Lemaire, 2015a). Furthermore, in applied numerical cognition, a decline in financial abilities as a subdomain of numerical ADL was reported (Finke et al., 2017). With domain-general functions globally declining with age (Li et al., 2004; Tucker-Drob et al., 2019), arithmetic deficits in the elderly have sometimes been explained as a consequence of domain-general – and not domain-specific – processes (Hinault & Lemaire, 2016a). However, systematic investigations of domain-general and domain-specific aspects across different arithmetic operations are still needed to comprehensively differentiate between age-related and pathological arithmetic deficits.

Arithmetic in Parkinson's Disease

In addition to non-pathological aging processes, neurodegeneration can further cause arithmetic impairments. Different dementias such as AD and PDD show both distinct cognitive profiles (Hugo & Ganguli, 2014) and common neuropathologies (e.g., the cholinergic system, Bohnen et al., 2003). Due to the lack of PD-specific studies on arithmetic, findings from AD research – although different pathophysiological mechanisms and predictive biomarkers are identified – might help to derive hypotheses regarding arithmetic deficits in PD. AD patients show deficits in complex written calculation, less flexibility in calculation strategies and higher error rates as well as longer response latencies in arithmetic (Arnaud et al., 2008; Lemaire & Leclère, 2014; Mantovan et al., 1999). Additionally, AD patients deteriorate when numerical ADL are concerned (Bangma et al., 2021; Martini et al., 2003; Sherod et al., 2009). This link between arithmetic skills and daily functioning is further supported by finding that the cortical volume of the angular gyri (involved in arithmetic fact retrieval) predicts financial deficits in MCI (Griffith et al., 2010). To conclude, the profile of cognitive impairment in AD is heterogeneous and the partial overlap with PDD neuropathology gives rise to limited, but informative conclusions for PD (Hugo & Ganguli, 2014).

The few studies addressing PD-specific arithmetic deficits identify relations between domain-general functions and deficits in different arithmetic operations. Arithmetic deficits in non-demented PD compared to a healthy control group were linked to working memory deficits (Liozidou et al., 2012) and executive dysfunction (Zamarian et al., 2006) for arithmetic word problems as well as complex arithmetic and calculation span tasks, respectively. However, Zamarian and colleagues (2006) did not find deficits in more basic arithmetic tasks (e.g., arithmetic fact retrieval, calculation span with low working memory load) and complex written calculation. Arithmetic performance in easier tasks might nevertheless be impaired in more advanced PD stages: Deficits in PDD have been reported in the basic arithmetic operations addition, subtraction, and multiplication in written arithmetic and all four operations in mental arithmetic (Kalbe, 1999a). Deficits in financial abilities, as one domain of numerical ADL were identified in PD-MCI and PDD patients (Martin et al., 2013). When analysing simple financial-arithmetic tasks in a sample of PD-NC and PD-MCI patients, there are first hints of impaired magnitude and place × value processing as well as deficits in the application of arithmetic procedures, shaped by gender, disease duration, attention and visuo-spatial/ constructional skills (Loenneker et al., 2021).

However, these studies are limited by several methodological constraints: All of them worked with small patient samples, diminishing the statistical power to detect small effect sizes as well as attenuating generalisability to the heterogeneous process of cognitive decline progression in PD. Due to the low number of items, difficulty varied unsystematically and the test-battery used by Zamarian and colleagues (2006) showed ceiling effects (Delazer et al., 2003). The four arithmetic operations have only

been investigated in PD-NC and PDD patients, but systematic evidence on PD-MCI is still lacking, questioning the onset of arithmetic deficits to be associated to distinctions in cognitive progression. To resume, available evidence on arithmetic in PD has neither systematically addressed the cognitive representations defining arithmetic deficits nor their association to cognitive deterioration progressing from PD-NC to PD-MCI.

Besides these behavioural findings, the neuroanatomical and -functional overlap of arithmetic processes (Klein et al., 2016) with progression of PD neuropathology (Caligiore et al., 2016) might also imply arithmetic deficits in PD. The sequential progress of PD neuropathology (classified by Braak stages one to six, Braak et al., 2003) affects brain areas needed for arithmetic, with deficits in domain-specific arithmetic areas suggesting processes comparable to primary acalculia.

First, number magnitude and place × value processing could be affected by degeneration in regions around the intraparietal sulcus, because of reduced acetylcholine production (caused by degeneration of the nucleus basalis magnocellularis beginning in Braak stage three) leading to a decrease in projections on the neocortex, such as parietal und temporal areas (beginning in Braak stage five; Jellinger, 2018; Moeller et al., 2015). Second, processing of visual number forms might be affected by degeneration of the temporal cortex (Arsalidou & Taylor, 2011; Koob et al., 2014). Lewy Bodies (LBs) in frontal, cingulate, and temporal cortex emerging in Braak stage five might further impact number processing (Arsalidou & Taylor, 2011; Collerton et al., 2003; Klein et al., 2016). Third, verbally mediated fact retrieval from memory in multiplication and division can be affected, both by a decreased basal ganglia gating function for arithmetic fact retrieval (Delazer et al., 2004; Owen et al., 1998) and affected semantic brain regions (i.e., inflammation in the angular gyrus, LB-induced degeneration of hippocampus and posterior cingulate gyrus; Jellinger, 2018; López González et al., 2016; Uribe et al., 2018). Fourth, addition and subtraction might also deteriorate as they require arithmetic fact retrieval and magnitude processing (Yang et al., 2017), and degeneration of the insula might further affect addition (Arsalidou & Taylor, 2011; Jellinger, 2018). Subtraction could be specifically affected by degeneration in premotor and supplementary motor areas (Braak et al., 2003, 2005; Kazui et al., 2000). Fifth, advanced mathematical problem solving can be affected by degeneration in semantic brain regions (angular, middle temporal, fusiform, inferior frontal, posterior cingulate, and parahippocampal gyri, dorsomedial and ventromedial prefrontal cortices, Zhou et al., 2018). To conclude, neuronal structures and circuits degenerating in PD and presumably being responsible for domain-specific numerical processes (such as magnitude processing, fact retrieval, calculation, and place × value integration) overlap. Whether or not these theoretical considerations will be empirically revealed as arithmetic deficits in PD is the aim of the current study.

However, examining domain-specific processes may not be sufficient. Globally, degeneration of domain-general functions supporting arithmetic processing can lead to secondary arithmetic deficits in PD. With degenerative processes in the dorsolateral prefrontal cortex associated with working memory, holding and monitoring task-relevant information should be impaired in complex arithmetic (Arsalidou & Taylor, 2011; Braak et al., 2003, 2005). The decrease of dopamine production in the striatum (beginning in Braak stage three) later reduces projections to the frontal lobe affecting processing speed, working memory, memory retrieval, verbal fluency, attention, and executive functions (Braak et al., 2003; Dirnberger et al., 2005; Martinez-Horta & Kulisevsky, 2019; Rinne et al., 2000). To conclude, PD-specific neurodegenerative processes suggest deficits in all four basic arithmetic operations, with severity depending on the stage of disease progression.

Objectives of the current study

Available studies on arithmetic in PD do not address the influence of cognitive status and the underlying cognitive representations of arithmetic. Therefore, the current exploratory study aims at investigating arithmetic deficits in PD patients with and without mild cognitive impairment and comparing between PD and a healthy control (HC) group. The influence of domain-general as well as domain-specific functions were differentiated to better understand the underlying mechanisms of behavioural arithmetic deficits. Finally, the probable use of arithmetic in the diagnostic of PD-immanent cognitive disorders was addressed by trying to discriminate the cognitive statuses of HC, PD-NC and PD-MCI using arithmetic tasks.

The research questions of the current study were:

Q1) Arithmetic deficits in PD: Are arithmetic performance and effects impaired in PD?

Q2) Influence of domain-general functions on arithmetic in PD: Do domain-general functions (e.g., visuo-spatial or executive functions) contribute to arithmetic performance in PD and if so for which operation?

Q3) Discrimination between cognitive statuses of PD by arithmetic performance: Can arithmetic performance be used to discriminate between a) PD-NC and PD-MCI and b) HC and PD-NC?

Arithmetic processes were systematically investigated concerning a) addition, b) subtraction, c) multiplication, and d) division. Task complexity was considered in terms of carry and borrow effects in addition and subtraction, respectively, and problem size effects in multiplication and division.

Whether and how PD-NC and PD-MCI differ in their arithmetic skills could not be inferred from the current literature. Therefore, we further explored the group effect by means of pairwise comparisons between HC and PD-NC and between PD-NC and PD-MCI to identify if arithmetic deficits are more frequent in patients than in controls and in patients with cognitive impairment than in those without. For

an overview of research questions, hypotheses and respective analysis plans with possible interpretations see Box 1.

Box 1. Study design

1) Are arithmetic performance and effects impaired in Parkinson's Disease?
Hypothesis (H1.1) There is a group effect in arithmetic performance: HC > PD-NC (H1.2) There is a group effect in arithmetic performance: PD-NC > PD-MCI **Analysis plan** Bayesian ANCOVAs with max. 6 clinical covariates <u>Addition:</u> 2 pairwise Bayesian mixed ANCOVAs: group (HC, PD-NC) × complexity (carry, non-carry) with clinical covariates on RT and ACC 2 pairwise Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (carry, non-carry) with clinical covariates on RT and ACC <u>Subtraction:</u> 2 pairwise Bayesian mixed ANCOVAs: group (HC, PD-NC) × complexity (borrow, non-borrow) with clinical covariates on RT and ACC 2 pairwise Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (borrow, non-borrow) with clinical covariates on RT and ACC <u>Multiplication:</u> 2 pairwise Bayesian mixed ANCOVAs: group (HC, PD-NC) × complexity (small, large problem size) with clinical covariates on RT and ACC 2 pairwise Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (small, large problem size) with clinical covariates on RT and ACC <u>Division:</u> 2 pairwise Bayesian mixed ANCOVAs: group (HC, PD-NC) × complexity (small, large problem size) with clinical covariates on RT and ACC 2 pairwise Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (small, large problem size) with clinical covariates on RT and ACC **Interpretation given different outcomes** (H1.1) Evidence for a group effect of HC > PD-NC: association of arithmetic deficits in PD with disease pathology, no mere aging effects. Evidence against group effect HC > PD-NC: arithmetic performance in PD not explained by disease-specific neurodegeneration, mere aging effect. (H1.2) Evidence for a group effect of PD-NC > PD-MCI: association of arithmetic deficits in PD with disease progression. Evidence against group effect PD-NC > PD-MCI: arithmetic performance in PD does not parallel global cognitive decline associated to disease progression. **Theory that could be shown wrong by the outcomes** Depending on results, neurodegeneration at stages of PD-NC and PD-MCI has or has not affected fronto-parietal circuits enough to impair arithmetic performance.
2) Do domain-general functions (e.g., visuo-spatial or executive functions) contribute to arithmetic performance in PD and if so for which operation?
Parallel hypotheses (H2.1) There is only a main effect of arithmetic complexity. (H2.2) There is only a main effect of cognitive covariate. (H2.3) There are main effects of cognitive covariate and arithmetic complexity with-/out interaction. **Analysis plan** Additional inclusion of one cognitive covariate per Bayesian ANCOVA <u>Addition:</u> 2 Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (carry, non-carry) with clinical and cognitive covariates on RT and ACC <u>Subtraction:</u> 2 Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (borrow, non-borrow) with clinical and cognitive covariates on RT and ACC <u>Multiplication:</u> 2 Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (small, large problem size) with clinical and cognitive covariates on RT and ACC

Division: 2 Bayesian mixed ANCOVAs: group (PD-NC, PD-MCI) × complexity (small, large problem size) with clinical and cognitive covariates on RT and ACC **Interpretation given different outcomes** (H2.1) Main effect of arithmetic complexity: after controlling for domain-general functions, arithmetic performance is driven by domain-specific functions. (H2.2) Main effect of cognitive covariate: domain-general function fully explains arithmetic performance. (H2.3) Main effects of cognitive covariate and arithmetic complexity with-/out interaction: joint contribution of domain-general and –specific functions to arithmetic. **Theory that could be shown wrong by the outcomes** Depending on the outcome, the results could suggest processes of arithmetic deficits in PD to be primary, secondary or indistinguishable.
3a) Can arithmetic performance be used to discriminate between PD-NC and PD-MCI? 3b) Can arithmetic performance be used to discriminate between HC and PD-NC, despite them showing comparable results regarding global cognition?
Hypothesis a) The weighted linear combination of the four basic arithmetic tasks can be used to differentiate whether a patient belongs to the group with or without mild cognitive impairments. b) The weighted linear combination of the four basic arithmetic tasks can be used to differentiate between HC and PD-NC (who show a comparable cognitive status). **Analysis plan** a) Bayesian logistic regressions with the predictors of z-standardized performance in addition, subtraction, multiplication, division and the covariates from research question 1 on cognitive status (PD-NC, PD-MCI) b) Bayesian logistic regression with the predictors of z-standardized performance in addition, subtraction, multiplication, division and the covariates from research question 1 on cognitive status (HC, PD-NC) **Interpretation given different outcomes** a1) If there is evidence for no effect of arithmetic, it is not suited to discriminate between PD-NC and PD-MCI, and arithmetic impairments seem to result from a different pathomechanism than global cognitive impairment. a2) If there is evidence for an effect of arithmetic, it is suited to discriminate between PD-NC and PD-MCI, and global cognitive and arithmetic impairments seem to share a common pathomechanism. b) Only if arithmetic performance differentiates HC from PD-NC, it can be used as an early marker for the detection of PD. **Theory that could be shown wrong by the outcomes** Depending on the results, degeneration of numerical cognition does (a2) or does not (a1) parallel global cognition in PD and is or is not only age-related as opposed to disease-specific (b).

Note. HC = healthy elderly without neurological impairments, PD = Parkinson's Disease, PD-NC = Parkinson's patients with normal cognition, PD-MCI = Parkinson's patients with mild cognitive impairment.

Methods

Statistical power analysis and sample size estimation

Empirical effect sizes were calculated (and transformed if required) with Psychometrica, JASP and the BF calculator for single-factor ANOVA summaries (Faulkenberry, 2019; JASP Team, 2021; Lenhard & Lenhard, 2016) to anticipate an approximate effect size for the current study (an overview of effect sizes can be found on https://osf.io/qgs5x/). Categorisation of Cohen's *d* as a measure of effect size follows Cohen (1992). Recent literature provides some evidence on arithmetic deficits in PD comparable to tasks used in the current study. Tamura, Kikuchi, Otsuki, Kitagawa, and Tashiro (2003) found a large difference ($d = 1.507$) between HC and non-demented PD in a calculation span of increasing length using single-digit numbers. The comparison of PD-NC patients and HC with the Graded Difficulty Arithmetic test (orally presented addition and subtraction tasks, increasing complexity, two- and three-digit operands, Jackson & Warrington, 1986) showed medium to large differences (d = 0.602, Scarpina et al., 2017; d =1.184, Zamarian et al., 2006). Orally presented word problems requiring mental calculation (WAIS arithmetic subtest, Wechsler, 1995) also yielded a large difference ($d > 0.8$) between non-demented PD patients and a HC group (Liozidou et al., 2012). Furthermore, Kalbe (1999a) showed a large difference ($d = 1.364$) in mental calculation between a PDD and HC group (all four basic arithmetic operations, single- and two-digit numbers).

As the smallest reported effect size for differences in arithmetic was medium to large, we expected an effect of at least $d = 0.5$ for the comparison between HC and PD-NC. Whether and how PD-NC and PD-MCI differ in their arithmetic skills could not be inferred from the current literature. Therefore, we further explored the group effect by means of pairwise comparisons between HC and PD-NC and between PD-NC and PD-MCI to examine if arithmetic deficits are more frequent in patients than in controls and in patients with cognitive impairment than in those without.

Following the procedure of sample size estimation in Bayes factor design analysis suggested by Schönbrodt and Wagenmakers (2018), participants were tested within a sequential Bayes factor design until the respective pairwise comparisons between HC and PD-NC and between PD-NC and PD-MCI reached a value of $BF_{10} \geq 6$ or $BF_{01} \geq 6$ for all of the four basic arithmetic operations in Q1 and Q2. The criterion for Q3 was based on the *BF* of including the respective predictor instead of excluding it ($BF_{inclusion}$) for each numerical predictor of the logistic regression on group. All three research questions needed to reach the criterion for the recruitment to stop. The repeated measures factors carry in addition and borrow in subtraction as well as the analysis of problem sizes in multiplication and division were conducted in an exploratory manner and therefore did not underlie considerations for sample size estimation. Due to feasibility, an additional maximum sample size of $n_{max} = 120$ valid data sets was

established (targeting equal group sizes). Which of these two criteria was reached first determined the ceasing of data collection. The first check of BF_{10} was planned for when all three groups had reached a size of 15 participants (i.e., $n_{min} = 45$), and checking was continued in steps of five additional participants per group. The process of sequential testing was additionally monitored with a Bayes factor plot indicating evidential development as a function of increasing sample size.

We estimated the properties of the planned research design with Monte Carlo simulations as implemented by Schönbrodt and Wagenmakers (2018) based on a sequential boundary of $BF_{10} = 6$, $d = 0.5$, $n_{min} = 35$, $n_{max} = 120$ and 10,000 simulated studies. Simulating the performance of our design under H_1 resulted in 9% of studies terminating at n_{max}, 90.8% terminating at H_1 boundary and 0.2% at H_0 boundary, on average stopping at $n = 63$. Simulating the performance of our design under H_0 resulted in 35.6% of studies terminating at n_{max}, 3.6% terminating at H_1 boundary and 60.7% at H_0 boundary, on average stopping at $n = 103$. After completion of data collection, the actual statistical power was assessed with a Bayesian power calculation for the effects of interest.

In case of participant exclusions, new participants were recruited for substitution. In case of early ceasing of the testing due to attrition, new participants were recruited, and data of the dropped out patient was only included in analyses when the patient gave informed consent and had already been assigned to a cognitive group based on the Movement Disorder Society (MDS) Task Force level I criteria (MoCA ≤ 26, Litvan et al., 2012).

Participants

This study has received approval by the ethics committee of the University of Tuebingen's medical faculty (161/2020BO2) and was registered not only at the Deutsches Register für Klinische Studien (DRKS-ID: DRKS00021091), but also at the World Health Organisation (Universal Trial Number: U1111-1257-2901). Patient recruitment was managed through the PD outpatient clinic in collaboration with rehabilitation facilities specialised in PD. Furthermore, PD patients who have been previously studied and gave consent (Ethical vote: 199/2011BO1) to be contacted for potential future study participation were contacted. The caregivers of the PD patients were also recruited as healthy controls, in accordance with defined inclusion and exclusion criteria, to make the HC group as comparable as possible considering sociodemographic background to both PD-NC and PD-MCI. Additionally, pensioners' initiatives were contacted for control group recruitment, with the study being advertised via the university mailing systems. All participants received monetary compensation (30€).

In the recruitment process, HC, PD-NC and PD-MCI were matched on the group level according to age ($M \pm 5$ years) and gender (max. 65% male) to approximately match sociodemographic and clinical

group means. This method cannot be used to correct for disease duration as cognitive status is confounded with disease duration in PD-MCI patients, who also have a longer history of PD, as indicated by PD-specific disease progression (S.-J. Lin et al., 2018). As PD medication can heavily influence experimental performance, all patients were tested in their "on-state" and were permitted to take their regular medication during the session if necessary.

The Level I MDS Task Force criteria for PD-MCI and PDD were used to assign patients to the PD-NC group or PD-MCI group, or excluded in the case of PDD (Emre et al., 2007; Litvan et al., 2012). This assumed a cut-off score of 26 on the Montréal Cognitive Assessment (MoCA, Nasreddine et al., 2005) to distinguish between PD-NC and PD-MCI; an absence of significant ADL impairment impacting everyday life excluded PDD. In order to acquire a cognitive diagnosis, assessment of the patients' self-impression, caregiver ratings, or ratings of the investigator was used to define slow progressive deterioration of cognition from a premorbid level. The HC group was assessed and excluded using the same MoCA cut-off score. The MoCA (Nasreddine et al., 2005; Thomann et al., 2018) is a short screening instrument for *global cognitive functioning,* evaluating short-term memory, visuo-spatial and executive functions, attention, language, and orientation to time and space. It has a maximum score of 30 and corrects for educational status in patients with 12 years of education or less.

Table 1 lists all inclusion and exclusion criteria separately for all three groups. The PD diagnosis was additionally confirmed by a movement disorder specialist in the outpatient clinic of the University Hospital of Tuebingen using the criteria of the United Kingdom Brain Bank (Hughes et al., 1992). This confirmation was required before inclusion into the current study and had to be documented in the patient's record. Additionally, in all cases, a relation of the PD patient (i.e., a spouse, child, friend, or relative of legal age) who has agreed to report on the presence and severity of PD related ADL problems (see section caregiver assessment) was asked to assess the patient.

Table 1. Inclusion and exclusion criteria by cognitive status.

(Sub-)group	Inclusion criteria	Exclusion criteria
General	• ≥ 60 years of age • (Corrected) hearing and vision • Fluent German skills & German schooling • Informed consent, voluntary participation • Stable health status (i.e., able to undergo the entire testing, comorbidities may be present but do not affect test performance) • Permission to ask a caregiver to verify patients rating on the occurrence of ADL problems	• Further neurological diseases impacting the central nervous system except for damage to intervertebral discs • (History of) substance abuse except for nicotine • BDI-II-Score ≥ 20 indicating major depression • Delirium or acute psychosis • Intake of anti-dementia drugs • History/ diagnosis of learning disabilities and developmental disorders (e.g., dyslexia, dyscalculia) • Known genetic diseases and family history of genetic diseases (min. 1 first or min. 2 second degree relatives) • Currently undergoing chemotherapy
HC	• Normal cognition (MoCA ≥ 26)	
PD	• Diagnosed idiopathic Parkinson's syndrome	• Severe impulse control disorder, dopamine dysregulation syndrome interacting with the patient's everyday life
PD-NC	• Normal cognition according to MDS Task Force level I diagnosis guidelines (e.g. MoCA ≥ 26)	• Diagnosis of cognitive impairment (level I-PD-MCI based on cognitive screening or PDD; Emre et al., 2007; PD-MCI or PDD, Litvan et al., 2012)
PD-MCI	• Mild cognitive impairment according to MDS Task Force level I diagnosis guidelines and diagnostic cut-off point based on Hoops et al. (Hoops et al., 2009, 26 > MoCA > 18)	• Diagnosis of PDD (Emre et al., 2007)

Note. Reprinted from "Deficits in or preservation of basic number processing in Parkinson's Disease? A registered report" by Loenneker et al., 2021, *Journal of Neuroscience Research.* Reprinted with permission.

Materials

Our materials and programmed experiments are openly available via PsychArchive (https://pasa.psycharchives.org/reviewonly/89feca909817dc421c3df8088ba8d597b7435dc8ac6fae36e7ed

d34b4b012bb7). All measures are reported in Table 2, except for the arithmetic tasks they are identical to those administered by Loenneker and colleagues (2024).

Introductory interview. To acquire sociodemographic (age, gender, years of education, handedness, and mother tongue) and clinically relevant (diagnosis, age at disease onset, medication for depression or cognition) information, an introductory interview was performed at the beginning of the experimental sessions. Furthermore, Levodopa equivalence dose (LEDD; DGN, 2016; Tomlinson et al., 2010) was calculated based on current dopaminergic medication and time since last dopaminergic medication intake was recorded. To gain insight about the patients' experience with mathematics, participants were asked to report preceding employment and associated mathematical experience (based on The International Standard Classification of Education ISCED 2011, UNESCO, 2012), level of income (below/ above/ average), math-related leisure activities, and self-rating of mathematical skill. For the cognitive diagnosis, patients were asked whether they have noticed a progressive deterioration of their cognition.

Motor performance. The severity of PD motor symptoms was determined via the sum score of the MDS revision of the Unified Parkinson's Disease Rating Scale Part III and IV (UPDRS III & IV, Goetz et al., 2008) and Hoehn & Yahr staging (Hoehn & Yahr, 1967). The motor examination of *UPDRS-III* comprised the domains speech, facial expression, rigidity of neck and extremities, hand and finger movements, toe tapping, leg agility, rising from a chair, (freezing of) gait, posture and stability, global spontaneity of movement, postural and kinetic tremor of hands, and rest tremor amplitude as well as constancy of rest tremor. The evaluation scale ranged from 0 = normal to 4 = severe, with more pronounced motor impairment being indicated by high values. A total score maximum of 132 based on 18 items (with several items relating to multiple extremities resulting in 33 scores) was calculated and included in the analysis. To assess dyskinesia, motor fluctuations, and dystonia, the subtest *UPDRS-IV* was employed, which is based on six items with a score ranging from zero to 24 points. As an additional measure for the severity of PD motor symptoms, the *Hoehn & Yahr score* ranging from 1 to 4 (1 = unilateral involvement only; 2 = bilateral involvement without impairment of balance; 3 = mild to moderate involvement, some postural instability, physical independence, and need for assistance in recovery from pull test; 4 = severe disability, and ability to stand and walk unassisted) was assessed.

Neuropsychological test battery. Performance in the cognitive domains of executive functions, working memory and attention, verbal memory, language, and visuo-spatial functions was assessed with at least one test per domain. Raw scores were considered as covariate measures for each test.

For *executive functions*, inhibition was assessed with the subtest Go/No Go "2 out of 5" in the computerized test battery of attention (TAP = Testbatterie zur Aufmerksamkeitsprüfung, Zimmermann &

Fimm, 2017). Participants had to discriminate five types of stimuli and respond to only two of them. The number of errors was used to measure the performance on this test.

Working memory abilities were assessed in a letter span forward and backward task (as in Artemenko et al., 2018). Participants listened to letter strings of increasing length. First, they had to reproduce the letter string in the same order, then reproduce it in reverse order. The maximum length of reproduced reverse letter strings was the relevant outcome variable. *Visuo-spatial working memory* was assessed using the Corsi Block-Tapping Test, in which the patients must mimic a sequence of blocks tapped on by the experimenter in the correct order, both forward and backward (Corsi, 1972; Kessels et al., 2000). Sequences increased in length, starting from two blocks to a maximum of nine, when both items of the same length were correctly reproduced. The relevant outcome measure was the longest Corsi backward span correctly reproduced. *Attention* was measured with the TAP subtest Alertness (Zimmermann & Fimm, 2017). The test required a simple reaction to a visually presented stimulus with or without prior notice via an auditory cue. The outcome variable assessed was the median RT in conditions without an auditory cue (as a measure of intrinsic arousal).

Verbal memory was measured with the German "Verbaler Lern- und Merkfähigkeitstest" (VLMT, Helmstaedter et al., 2001). Participants had to learn, recall, and recognize super-span lists of words after different time intervals while coping with an intruding list of distractors. The relevant measure of performance was the sum score for delayed recall.

Language was measured with the German version of the WAIS-IV subtest Similarities (Petermann, 2012), operationalised as conceptual understanding, which were assessed via an indication of what two terms have in common. The total number of correctly solved items was used as the outcome measure.

Visuo-spatial function was assessed using the Benton Line Orientation Test (Benton et al., 1978), in which two lines of a certain angle and position were presented on a sheet of paper. Participants had to compare these two lines with eleven lines being arranged in the shape of a star around a centre. Correctly identifying the two lines from the displayed eleven lines was scored as one point, with a maximum sum score of 15.

Arithmetic tasks. Arithmetic performance was assessed in an oral production paradigm by computerised tasks of the four basic arithmetic operations programmed with OpenSesame (Mathôt et al., 2012). Experimental trials were preceded by practice items, which could be repeated if participants did not understand the instruction in the first attempt. Practice trials consisted of easier tasks (e.g., using the operand 1 and single-digit numbers). There were no time restrictions for solving the arithmetic problems, but testing was stopped if a participant could not answer any of the first ten experimental trials of the current task. All stimuli were presented in black against a white background and in randomised order

within the current basic arithmetic operation. Every trial started with a fixation point in the shape of "o" for 750 ms and was followed by the arithmetic problem presented centrally on the screen until the participant pressed the space key while responding orally. The participant started pressing the space key when starting to answer, held it in the meantime and released the key when having finished answering. The critical RT here was the first key press, the time between key press and key release was used to exclude participants taking too long due to utterances intermixed with the answer. The numerical response was entered with a QWERTZ keyboard by the experimenter, who then initiated the next trial (for a similar procedure in elderly see Artemenko, 2021). By decreasing motor effort for PD patients, this response format aimed at minimizing the influence of PD-immanent motor impairments on arithmetic performance while logging both response and RT. Outcome measures were both proportion of correctly solved trials (ACC) and RT for each basic arithmetic operation. The factor of item complexity was operationalized depending on the basic arithmetic operation: carry operation for addition, borrow operation for subtraction, and problem size for multiplication and division.

Addition. The 50 experimental trials were preceded by five practice trials. All addition problems consisted of two two-digit numbers (with results ranging from 36 to 96). Items did not imply pure decades (e.g., 20), ties (e.g., 22), unit ties (e.g., 32 + 52), or decade ties (e.g., 23 + 25). Carry and non-carry items were matched considering problem size, numerical size of both operands and units and decades of both operands, parity of both operands, and position of the smaller operand. This allowed for calculating the carry effect, defined as the difference in mean RTs between carry and non-carry problems, counterbalanced for problem size (which was defined as the sum of the two operands). Therefore, outcome measures for the task-based approach were RT and ACC, whereas the outcome measure for the effect-based approach was the carry effect (see Artemenko et al., 2018).

Subtraction. The 50 experimental trials were preceded by five practice trials, which all consisted of two two-digit numbers. The subtraction problems were constructed as the inverse problems of addition (e.g., 32 + 25 → 57 - 25). This allowed for calculating the borrow effect, defined as the difference in mean RTs between borrow and non-borrow problems, counterbalanced for problem size (which is defined as the value of the first operand, i.e., the minuend). Therefore, outcome measures for the task-based approach were RT and ACC, whereas the outcome measure for the effect-based approach was the borrow effect (see Artemenko et al., 2018).

Multiplication. The 45 experimental trials were preceded by five practice trials. All multiplication problems came from single-digit multiplication tables (numbers 1-9), with single- and two-digit results ranging from 1 to 81. The item set included each number pair only once (e.g., 5 × 3 or 3 × 5) with the position of the larger operand being counterbalanced, and ties (e.g., 5 × 5). Items consisted of small

(product ≤ 25) and large (product > 25) problem sizes (as in Archambeau et al., 2019; Grabner et al., 2009). The outcome measures were both overall ACC and RT.

Division. The 45 trials were preceded by five practice trials. The division problems were constructed as the inverse problems of multiplication (e.g., e.g., 5 × 3 → 15 ÷ 5). Items consisted of small (divisor ≤ 25) and large (divisor > 25) problem sizes (being the inverse of the definition of problem size for multiplication). The outcome measures were both overall ACC and RT.

Clinical questionnaires for non-motor assessment, ADL and quality of life. In addition to the introductory interview, self-report questionnaires were used to assess other clinical variables. The Nonmotor Symptoms Questionnaire for Parkinson's Disease (NMSQuest; Chaudhuri et al., 2006) is a 30-item screening questionnaire with a categorical answering format of "yes", "no", and "don't know" based on the occurrence of *non-motor symptoms* in the last month. It allows for the quantification of sleep disorders and neuropsychiatric, autonomic, gastrointestinal, sensory, and other symptoms by means of a sum score.

The Beck Depression Inventory (BDI-II; Hautzinger et al., 2006) was used to assess the patients' *depressive symptoms*, with a cut-off of 20, which indicates signs for major depression. The assessment consists of 21 items which the patients must rate for the intensity of occurrence of the symptom during the last two weeks on a scale ranging from 0 to 3 (maximum sum score of 63).

The patients' current *activities of daily living* were measured using the Functional Activities Questionnaire (FAQ; Pfeffer et al., 1982), which is tailored to older adults. Participants had to rate their level of performance (ranging from 0 = normal to 3 = dependent) in 10 different daily life activities, resulting in a sum score characterising ADL.

Health-related quality of life was assessed with the single index score of the 39-item Parkinson's Disease Questionnaire (PDQ-39; Jenkinson et al., 1997). The eight dimensions such as activities of daily living and social support were coded on a scale ranging from 0 (= perfect health) to 100 (= worst health).

Caregiver assessment. For a broader evaluation of the participant's ADL, the FAQ was also completed by a caregiver. Additionally, caregivers were asked for how they are related to the participant, their demographic data (i.e., age, gender), and how frequently and intensely they were in contact with the participant. To assess the cognitive diagnosis, caregivers were asked whether they had noticed a progressive deterioration of the patient's cognitive state.

Table 2. Summary of collected measures

Measure	Aspects	Description/ Example	Outcome
Introductory interview	Sociodemographic information	Age, gender, years of education, handedness, mother tongue, experience with mathematics, level of income	
	Clinical information	Diagnosis, age at disease onset, medication for depression/ cognition, LEDD	
Motor performance	UPDRS-III	Speech, facial expression, rigidity of neck / extremities, hand / finger movements, toe tapping, leg agility, rising from a chair, (freezing of) gait, posture, stability, global spontaneity of movement, postural / kinetic tremor of hands, rest tremor amplitude / constancy	Total score (max. 132 = pronounced impairment)
	UPDRS-IV	Dyskinesia, motor fluctuations, dystonia	Total score (max. 24 = pronounced impairment)
	Hoehn & Yahr staging	Uni-/ bilateral involvement, postural instability	Score (1 – 4)
Neuropsychological test battery	TAP Go/ No Go	Executive function (inhibition)	*N* of errors
	MoCA	Global cognition	Sum score (max. 30 = no impairment)
	Letter span backward	Verbal working memory	Max. span
	Corsi Block-Tapping Test backward	Visuo-spatial working memory	Max. span
	TAP Alertness	Attention	Median RT without cue
	VLMT	Verbal memory	Sum score delayed recall
	WAIS-IV similarities	Language	*N* correct
	Benton Line Orientation Test	Visuo-spatial function	*N* correct
Arithmetic tasks	Addition	5 practice, 50 experimental (50% carry operations), 2-digit numbers	RT, ACC, carry effect
	Subtraction	5 practice, 50 experimental (50% borrow operations), 2-digit numbers	RT, ACC, borrow effect
	Multiplication	5 practice, 45 experimental (50% small problem size), 1-digit numbers for multiplication (inverse for division)	RT, ACC, problem size effect
	Division		
Clinical questionnaires	NMSQuest	Nonmotor Parkinson's symptoms	Sum score (max. 30)
	BDI-II	Depressive Symptoms	Sum score (max. 63 =

			pronounced impairment)
	FAQ	Activities of daily living	Sum score (max. 30 = pronounced impairment)
	PDQ-39	Health-related quality of life	Sum score (0 = perfect health to 100 = worst health)

Procedure

As part of a broader research project on numerical cognition in PD, the current study was conducted in joint sessions with the registered report that investigates basic number processing in PD, which had already been granted in-principle acceptance before in-principle acceptance of the current registered report (for further measures conducted regarding basic number processing in PD see Loenneker et al., 2024). After obtaining written informed consent, the predefined inclusion and exclusion criteria were used to assess participant eligibility in a semi-standardised questionnaire. PD patients' cognitive performance was used to assign them to the PD-NC or the PD-MCI group. Participants attended two sessions of 1.5 to 2 hours each. In order to handle patient attrition, there were breaks within each session as required. The first experimental session consisted of the sociodemographic questionnaire, the basic numerical tasks (not considered here: transcoding, number line estimation, non-symbolic magnitude comparison, symbolic magnitude comparison, Loenneker et al., 2024) and the arithmetic tasks addition, subtraction, multiplication and division in this order. Afterwards, participants and caregivers completed the clinical scales and questionnaires which could also be filled out at home between the first and second session. In the second session, the MoCA, clinical variables, motor assessment, and neuropsychological test battery were conducted in that order. The two sessions could be scheduled three weeks apart at a maximum. On a conceptual level, the first publication (Loenneker et al., 2024) addressed the basic foundations of number processing, whereas the current publication focuses on arithmetic skills, which are more relevant for a patient's daily life. Piloting the entire experimental procedure with two healthy elderly persons did not lead to any changes regarding the measures reported in the current article.

Data treatment and proposed analysis pipeline

Data analyses were run with R version 4.2.2 (R core team, 2014) and JASP version 0.14.1 (JASP Team, 2021). Data was managed using REDCap electronic data capture tools (Harris et al., 2009). Anonymised data and analyses scripts are freely available on PsychArchive (https://pasa.psycharchives.org/reviewonly/89feca909817dc421c3df8088ba8d597b7435dc8ac6fae36e7ed

d34b4b012bb7). As all analyses were conducted within the framework of Bayesian statistics, correction for multiple comparisons were not necessary (as elaborated by Gelman et al., 2012). For all inferential statistics an α-level of .05 was assumed and effect sizes were estimated. Classification of the effect size Cohen's *d* was based on Cohen (1992) and Bayes Factors (*BF*) followed Jeffreys (1961). BF_{10} indicated the probability of evidence in favour of the alternative hypothesis and BF_{01} indicated the probability of evidence in favour of the null hypothesis with $BF_{01} = 1/BF_{10}$. Additionally to reporting BF_{10}, we reported how many participants matched theoretical expectation as Percent Correct Classification (PCC) following Grice and colleagues (2020). Overall, planning, analysis, interpretation and reporting of results followed recommendations by van Doorn et al. (2020).

Data preprocessing

Exclusions. In case of missing data, participants were excluded in a case-wise manner for those analyses including the respective measure. Participants needed to achieve an ACC of minimum 75% in tasks in a forced choice format (TAP tasks, Benton line orientation test) for inclusion in the respective analysis. The arithmetic tasks began with practice trials and could be aborted if none of the first ten experimental trials was solved correctly. In this case, we assigned a value of 0 for ACC and the participant was excluded from the RT analysis but included in the ACC analysis. Participants whose performance exceeded 3 *SD* below the group *M* were excluded from the respective analysis.

Reaction times. Data trimming for RTs in the arithmetic tasks was adapted from Baayen & Milin (2010). The RT distribution of correctly solved trials was inspected with by-subject quantile-quantile plots, in order to identify the best suited theoretical model for data transformation to approach normally distributed data. The most accurate distribution was determined with the best model fit (see Suppplementary Material B, Table S1 for the respective distributions) and used to transform the data. After that, outliers were excluded in two steps. First, anticipations were excluded defined as RTs faster than 200 ms. Second, model criticism was used to exclude remaining outliers, based on Shapiro tests for normality. Those data points with absolute standardized residuals exceeding 3 were removed. It was further preregistered that, temporal dependencies should be corrected for with autocorrelation functions by participant and a regression model fitted to responses with a log-transformation for latencies and including the covariates trial number and preceding RT. However, this step could not be conducted, as it is incompatible with hypothesis testing using ANCOVAs which are running on averaged data (only appropriate for generalized linear models considering each observation of each subject; see Supplementary Material A). An overview of chosen transformations and excluded observations per preprocessing step can be found in the Supplementary Material B (Table S1). For participants to be included in the RT analysis, a minimum of five valid data points out of 20-25 trials needs to be available

per condition (i.e., minimum ACC of 20-25%).**Accuracy.** Either arcsine- or logit-transformation were chosen for ACC data depending on the best model fit for the empirical distribution of ACC.

Assumption check. Following the aforementioned phase of data pre-processing, assumptions for respective statistical hypothesis testing were checked. Assumptions were tested by means of visual inspection (scatterplots, residual plots, residual boxplots), frequentist (Levene test, Mauchly's test) and Bayesian (variance homogeneity, Dablander et al., 2020) tests. In case of a violation of assumptions for the ANCOVA regarding normal distribution and variance homogeneity of residuals, appropriate transformations were conducted. Predictors of the logistic regression were checked for collinearities based on a variance inflation factor below 10.

Group-wise characteristics

Possible variables confounding the experimental manipulations were identified as differences in sociodemographic, clinical and cognitive variables between the three groups. The categorical variables gender and Hoehn & Yahr stage were characterized as total number per category and corresponding percentage, and compared with Bayesian contingency tables between HC and PD-NC and between PD-NC and PD-MCI. Continuous variables were described with *M* (SD) and compared between HC and PD-NC and between PD-NC and PD-MCI with Bayesian independent samples *t*-tests. The candidate variables to be considered as confounders were sociodemographic (age, education years, level of income, educational and professional math experience) and clinical (disease duration, age at onset, LEDD, intake of antidepressants) variables. Furthermore, motor (UPDRS-III & IV) and cognitive function were compared. Last, group-wise comparisons were conducted on clinical questionnaires (non-motor symptoms: NMSQuest, depression: BDI-II, ADL self-report & caregiver report: FAQ, health-related quality of life: PDQ-39). Additionally, irrespective of the status of mild cognitive impairment, the number of HC, PD-NC, and PD-MCI participants with impairments in any of the cognitive measures were reported.

Hypothesis testing

We expected to find differences of at least medium to large effect sizes between the HC and PD-NC and between the PD-NC and PD-MCI group. For all Bayesian analyses, Cauchy priors centred on zero with a scale factor of 1 were used based on findings of Zamarian et al. (2006), and models were compared to the null model.

We estimated the robustness of the BF analysis based on Zamarian's (2006) results on the Graded Difficulty Arithmetic Test (i.e., mixed arithmetic tasks), compared between PD-NC (M = 10.2, SD = 4.3, n = 15) and HC (M = 15.5, SD = 4.6, n = 28), $t(41)$= -3.70, $p < .001$ ($M_{difference}$ = -5.3 raw units, SD_{pooled} =

4.49 raw units, SE_{pooled} = 1.41 raw units). We assessed the robustness using JASP (JASP Team, 2021). As this difference of mean accuracies reflects an effect size of d = 1.18, we used a scale factor of 1 for the prior Cauchy distribution. We defined the likelihood based on a non-central d distribution with the parameters from Zamarian's results ($M_{difference}$ = 1.18 d units, SE_d = 0.34 d units, observations = 43), assuming a one-sided Cauchy distribution (location = 0, scale = 1) with a lower limit at 0 for the model of the alternative hypothesis, and a distribution with a spike at 0 for the null hypothesis. Results on possible ranges of the scale factor producing BFs indicating conclusive evidence are reported in Table 3. However, the study by Zamarian et al. (2006), that we base our effect size estimates on, differs in important aspects from our planned study: (1) they report accuracies instead of RT (we use both separately), (2) they compare a healthy control group with a PD group without cognitive impairments (we additionally compare this group with a patient group with cognitive impairments), and (3) they use a task-based instead of our effect-based approach. While findings from Zamarian et al. (2006) provide a first hint in which region the expected effect sizes and priors could be, we are quite far-off a direct replication: Hence, our effect sizes could differ considerably from Zamarian et al. (2006). Considerations regarding effect sizes in Q3b cannot be inferred from the literature, as we are not aware of a study comparing PD-NC with PD-MCI patients in mental arithmetic.

Table 3. Robustness considerations of scale factors (in d units) for Bayesian analysis.

Model for alternative hypothesis		
Location	Scale	BF_{10}
0	1	19,370,000
0	0.1	3,530,000
0	0.01	356,145.75
0	0.001	35618.52
0	0.0001	3561.87

Table 3 indicated that we could expect fairly robust results up to very small scale factors. Since the current study implied both comparisons with a PD-NC and a more advanced PD-MCI group as opposed to a single PD-NC sample in Zamarian's study, we might have even expected larger effects. However, we wanted to be careful with this prediction as target tasks, control variables, and items within tasks differ and may modulate effects. After data acquisition, robustness of the BF across different scale factors was assessed and can be found in the respective ANCOVA tables.

Zamarian's study also provided evidence regarding the association of domain-general factors with arithmetic performance. Where they found a difference in complex mental calculation (GDAE) between patients (M = 10.2, SD = 4.3) and controls (M = 15.5, SD = 4.6), t(41)= -3.70, p < .001, they identified associations between the GDAE and interference naming (r = - 0.633, p <0.02), digit span

forward (r =0.625, p <0.02) and block span backward (r =0.584, p <0.03). These correlations can be transformed into effect sizes of d = -1.64, d = 1.60, and d = 1.44, respectively.

Manipulation check. To ensure the quality of our data, we analysed the carry effect for addition, the borrow effect for subtraction, and the problem size effect for multiplication and division in the HC group for ACC and RT to see whether we find the usual arithmetic effects in our data. This was done with Bayesian t-tests.

Arithmetic deficits in PD (Q1). The three groups were first compared regarding sociodemographic and clinical variables. We assumed age, gender, education, level of income, educational and professional math experience, Hoehn & Yahr staging, disease duration, and depression could significantly differ between groups. In this case we included a maximum number of six clinical covariates in all models. ANCOVAs were run with the respective previously identified covariates. The factor group was split among two chains of separate pairwise analyses, comparing HC with PD-NC and PD-NC with PD-MCI. *Addition* was analysed by two Bayesian mixed ANCOVAs with the factors group (HC vs. PD-NC/ PD-NC vs. PD-MCI) and complexity (non-carry, carry) on RT and ACC. *Subtraction* was analysed by two Bayesian mixed ANCOVAs with the factors group (HC vs. PD-NC/ PD-NC vs. PD-MCI) and complexity (non-borrow, borrow) on RT and ACC. *Multiplication* was analysed by two Bayesian mixed ANCOVAs with the factors group (HC vs. PD-NC/ PD-NC vs. PD-MCI) and complexity (small, large problem size) on RT and ACC. *Division* was analysed by two Bayesian mixed ANCOVAs with the factors group (HC vs. PD-NC/ PD-NC vs. PD-MCI) and complexity (small, large problem size) on RT and ACC. (H1) We expected the HC group to perform better than the PD-NC and PD-MCI groups, and the PD-NC group than the PD-MCI group. The group comparisons were conducted by pairwise ANCOVAs that include the covariates of the respective analysis.

Influence of domain-general functions on arithmetic processing in PD (Q2). The question whether arithmetic performance within the group of PD patients can be explained by domain-general functions was addressed with further Bayesian ANCOVAs. These analyses were preceded by Bayesian correlational analyses of arithmetic performance with possible covariates. For each arithmetic task, the cognitive covariate with the highest association (minimum significant correlation of 0.3) was included in the respective model, adding up to a maximum of seven covariates including confounders identified for Q1. Bayesian ANCOVAs conducted for Q1 were then repeated for Q2 with the respective additional cognitive covariate and only the two groups PD-NC and PD-MCI. There were three possible outcomes: (H2.1) A main effect of arithmetic complexity indicates that after controlling for domain-general functions, arithmetic performance is driven by domain-specific functions. (H2.2) A main effect of cognitive covariate indicates that the domain-general function fully explains arithmetic performance. (H2.3) There is both a main effect of the respective cognitive covariate and arithmetic complexity

with(out) an interaction effect of these two, hinting at a joint contribution of domain-general and –specific functions to performance in the respective arithmetic task.

Discrimination between cognitive statuses of PD patients by arithmetic performance (Q3). The last question targeting the diagnostic use of arithmetic was answered using two Bayesian logistic regressions. This discriminant analysis was conducted with z-standardized performance in addition, subtraction, multiplication, and division as multiple predictors and with the dependent variable of cognitive status for a) PD-NC and PD-MCI and b) HC and PD-NC. Covariates from Q1 were also included in the model. Both influential case diagnostics and outlier analysis were applied to minimize the effect of highly influential participants on the regression. The probability for each person to fall into the respective group was calculated based on the regression. We hypothesized that the respective group could be predicted with a linear combination of the arithmetic predictors. H3a) The combined performance of the four arithmetic tasks can be used to differentiate whether a patient belongs to the group with or without mild cognitive impairments. H3b) The combined performance of the four arithmetic tasks can be used to differentiate between HC and PD-NC, while these groups show a comparable cognitive status. There were several possible unexpected outcomes: (H3a1) If there is evidence for no effect of arithmetic, it is not suited to discriminate between PD-NC and PD-MCI and arithmetic impairments seem to result from a different pathomechanism than cognitive impairment. (H3a2) If there is evidence for an effect of arithmetic, it is suited to discriminate between PD-NC and PD-MCI and cognitive and arithmetic impairments seem to at least partly share a common pathomechanism. (H3b) If arithmetic performance only differentiates HC from PD-NC, but not PD-NC from PD-MCI, it could be used as an early marker for the detection of PD.

Results

Sample descriptives

Overall, $N = 150$ participants were tested, of which $N = 119$ ($n = 40$ HC, $n = 50$ PD-NC, $n = 29$ PD-MCI) were included in the analyses. Poor performance in the MoCA was the main exclusion criterion for patients (score < 19, $n = 3$) and HC (score < 26, $n = 20$), followed by comorbid psychiatric disease in patients (major depression, BDI-II > 19, $n = 6$) and HC ($n = 1$). Further exclusion criteria in the PD patient group included the presence of other neurological diagnoses (stroke, $n = 1$), intake of dementia medication ($n = 3$), and drop-out for medical reasons ($n = 1$). For some participants, multiple exclusion criteria applied. Deviations between stage 1 and stage 2 of the registered report are summarized in Supplementary Material A.

The current article reports on the same sample described in Loenneker et al. (2024). The PD-MCI group was smaller than the other two groups, as their recruitment proved most challenging due to the two long testing sessions. The HC and PD-NC groups differed in years of education, severity of motor symptoms (UPDRS-III), and depression (BDI-II), while the PD-NC and PD-MCI groups differed in levodopa equivalence dose (LEDD) and Hoehn & Yahr stage. These confounders were therefore included in the respective pairwise analyses (see Table 4). Descriptive information on group performance in numerical tasks (Table S2), cognitive tasks (Table S3), and math-related characteristics (Table S4) can be found in the Supplementary Material C.

Table 4.

Sociodemographic and clinical characteristics per group.

	HC (n = 40)	PD-NC (n = 50)	HC vs. PD-NC BF_{10}	PD-MCI (n = 29)	PD-NC vs. PD-MCI BF_{10}
Sociodemographic					
Age *M (SD)*	68.8 (5.80)	69.6 (6.11)	0.27	70.9 (5.82)	0.35
Gender (female) *n* (%)	20 (50.0%)	22 (44.0%)	0.44	7 (24.1%)	1.74
Education years *M (SD)*	17.0 (2.92)	15.3 (3.75)	**3.01**	13.3 (3.58)	2.33
Clinical					
Disease duration *M (SD)*		7.45 (4.88)		9.27 (5.22)	0.68
Age at onset *M (SD)*		62.1 (7.37)		61.6 (6.38)	0.26
Levodopa equivalence dose *M (SD)*	0*	648 (435)		997 (656)	**7.18**
Antidepressants *n* (%)	0*	8 (16.0%)		3 (10.3%)	0.33
Motor symptoms (UPDRS-III) *M (SD)*	1.30 (1.54)	26.3 (12.7)	**4.48**	32.7 (11.0)	2.21
Hoehn & Yahr *n* (%)					**4.44**
0		0		0	
1		11 (22.0%)		0 (0%)	
2		31 (62.0%)		21 (72.4%)	
3		7 (14.0%)		6 (20.7%)	
4		1 (2.0%)		2 (6.9%)	
Motor type *n* (%)					0.36
Tremor		17 (34.0%)		6 (20.7%)	
PIGD		26 (52.0%)		19 (65.5%)	
mixed		7 (14.0%)		3 (10.3%)	
Depression (BDI-II) *M (SD)*	3.08 (3.78)	7.83 (5.12)	**> 1000**	8.71 (4.71)	0.31

Notes. Two-sided Bayesian independent samples *t*-tests with $r = 0.707$ and JZS prior distribution for continuous variables and Bayesian contingency tables with joint multinomial sampling for categorical variables. $BF_{10} > 3$ are marked in bold. * Intake of both antidepressants and dopaminergic medication is an exclusion criterion for the control group.

Research question 1) Are there group differences in arithmetic between healthy controls, PD patients without cognitive impairment and PD patients with mild cognitive impairment?

The complete Bayesian ANCOVA tables concerning research question Q1 can be found in Supplementary Material D, while results regarding the most important parameters including robustness checks are reported in the main text. ANCOVA results addressing research question 1 are summarized in Table 5 and Figure 1. Evidence accumulation during sequential testing is displayed in Supplementary Material G, Figures S2 to S5, which indicates instable evidence for addition, and stable evidence for the other three arithmetic operations.

Table 5

Bayesian ANCOVAs on addition, subtraction, multiplication and division performance comparing HC, PD-NC, and PD-MCI groups

		A) Addition			B) Subtraction			C) Multiplication			D) Division		
		BF_{incl} for *r* scale			BF_{incl} for *r* scale			BF_{incl} for *r* scale			BF_{incl} for *r* scale		
I) Accuracy		1	0.5	2	1	0.5	2	1	0.5	2	1	0.5	2
a) HC vs. PD-NC	Participant	*0.10*	*0.10*	*0.12*	**> 1000**	**> 1000**	**> 1000**	*0.06*	*0.06*	*0.06*	1.96	1.80	2.04
	Education	0.45	0.36	*0.29*[#]	0.87	0.91	0.92	0.69	0.65	0.70	1.05	1.03	1.02
	Motor symptoms	0.36	0.38	0.39	0.53	0.50	0.54	0.91	0.85	0.96	**3.37**	**3.29**	**3.59**
	Depression	0.34	0.45	0.34	0.38	0.35	0.41	0.40	0.39	0.43	*0.30*	*0.30*	*0.31*
	Group	0.96	1.70	0.63	1.54	2.02	0.81	*0.18*	0.35[#]	*0.09*	*0.20*	0.35[#]	*0.11*
	Complexity	**172.69**	**222.27**	**85.27**	**34.91**	**50.33**	**20.49**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**
	Group × complexity	*0.17*	0.34[#]	*0.10*	0.39	0.60	*0.19*[#]	*0.14*	*0.28*	*0.08*	*0.12*	*0.23*	*0.06*
b) PD-NC vs. PD-MCI	Participant	*0.04*	*0.03*	*0.04*	**> 1000**	**> 1000**	**> 1000**	*0.26*	*0.23*	*0.27*	0.91	0.73	0.72
	Levodopa equivalence dose	*0.20*	*0.19*	*0.20*	0.53	0.54	0.53	*0.22*	*0.22*	*0.22*	0.35	0.39	0.40
	Disease stage	*0.26*	*0.24*	*0.24*	0.37	0.36	0.38	0.81	0.77	0.79	**7.79**	**6.44**	**7.31**
	Group	**66.69**	**94.18**	**38.38**	**15.14**	**16.59**	**9.98**	*0.12*	*0.24*	*0.07*	*0.16*	*0.32*	*0.09*
	Complexity	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**53.39**	**69.00**	**27.76**
	Group × complexity	*0.25*	0.41[#]	*0.13*	*0.31*	0.43[#]	*0.14*	*0.13*	*0.23*	*0.07*	*0.28*	0.45[#]	*0.13*
II) Reaction Time													
a) HC vs. PD-NC	Participant	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**
	Education	**3.38**	1.87[#]	**3.32**	**21.48**	**23.44**	**20.02**	**3.12**	**3.65**	**3.68**	**43.65**	**54.79**	**38.85**
	Motor symptoms	**5.93**	**3.50**	**3.77**	**3.04**	**3.63**	1.91[#]	**10.92**	**13.07**	**7.46**	**113.61**	**195.63**	**74.10**
	Depression	*0.16*	*0.26*	0.65[#]	0.33	*0.32*[#]	0.33	0.31	0.29	0.33	0.24	0.20	0.22
	Group	2.25	2.71	**4.08**[#]	**61.15**	**37.74**	**74.24**	**5.74**	**5.93**	**3.28**	**11.95**	**13.80**	**6.47**
	Complexity	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**
	Group × complexity	0.94	1.63	0.52	*0.03*	*0.09*	*0.01*	*0.08*	*0.25*	*0.02*	*0.03*	*0.10*	*0.01*
b) PD-NC vs. PD-MCI	Participant	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**
	Levodopa equivalence dose	0.33	0.27	0.28	0.34	0.35	0.38	*0.22*	*0.12*	*0.24*	*0.19*	*0.20*	*0.17*
	Disease stage	**45.50**	**57.64**	**28.45**	**30.88**	**38.02**	**18.68**	**147.90**	**345.14**	**131.80**	**> 1000**	**> 1000**	**838.58**
	Group	1.60	0.77	0.95	**38.94**	**23.24**	**34.41**	**12.85**	**27.75**	**9.05**	**56.93**	**47.15**	**38.73**
	Complexity	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**	**> 1000**
	Group × complexity	0.78	1.86	*0.26*[#]	*0.03*	*0.11*	*0.01*	*0.05*	*0.10*	*0.02*	*0.03*	*0.11*	*0.01*

Notes. Models were compared to the null model and effects were estimated across matched models. Group and complexity were treated as fixed factors, while subjects were treated as random factor. [#] The robustness check yielded a deviating result. BFs > 3 were marked in bold and provide evidence in favour of including the respective variable. BFs < 1/3 were marked in italics and provide evidence against including the respective variable. BFs that were not marked should be interpreted as inconclusive. Evidence in favour of group effects is shaded in grey.

1a) Addition

As a quality check, Bayesian paired *t*-tests were conducted in the HC group to test the complexity effect: evidence was inconclusive (BF_{10} = 0.59) for accuracy differences between non-carry (*M* = 0.98, *SD* = 0.03) and carry (*M* = 0.96, *SD* = 0.04) problems, but there was evidence for a reaction time difference (BF_{10} = 3.92), with faster calculation for non-carry (*M* = 2847 ms, *SD* = 1110 ms) than for carry (*M* = 3890 ms, *SD* = 1928 ms) problems. Moreover, ANCOVAs for both outcome measures indicated substantial evidence for the effect of complexity (see Table 5A).

To compare *PD-NC* to *HC*, Bayesian ANCOVAs were performed on accuracy and reaction time in addition, including group (HC, PD-NC) and complexity (carry, non-carry) as fixed factors, and education, UPDRS-III, and depression as covariates (see Table 5A and Figure 1A). For *accuracy* (analysis A.I.a), the model including complexity was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, while there was inconclusive evidence regarding group and evidence against the interaction of group with complexity. For *reaction time* (analysis A.II.a), the model including group, complexity, participant, education, and depression was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity and participant, while there was inconclusive evidence regarding group and the interaction of group with complexity.

To compare *PD-MCI* to *PD-NC*, Bayesian ANCOVAs were performed on accuracy and reaction time in addition, including group (PD-NC, PD-MCI) and complexity (carry, non-carry) as factors and LEDD and Hoehn & Yahr stage as covariates (see Table 5A and Figure 1B). For *accuracy* (analysis A.I.b), the model including group and complexity was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity and group (PD-NC being more accurate than PD-MCI), and evidence against the interaction of group with complexity. For *reaction time* (analysis A.II.b), the model including group, complexity, participant, and Hoehn & Yahr disease stage was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant, and Hoehn & Yahr disease stage, while there was inconclusive evidence regarding group and the interaction of group with complexity.

1b) Subtraction

As a quality check, Bayesian paired *t*-tests were conducted in the HC group to test the complexity effect: there was evidence against accuracy differences (BF_{10} = 0.25) between non-borrow (*M* = 0.94, *SD* = 0.06) and borrow (*M* = 0.93, *SD* = 0.06) problems, but there was evidence for a reaction time difference (BF_{10} > 1000), with faster calculation for non-borrow (*M* = 4190 ms, *SD* = 1862 ms) than for borrow (*M*

= 5502 ms, *SD* = 2289 ms) problems. Moreover, ANCOVAs for both outcome measures indicated substantial evidence for the effect of complexity.

To compare *PD-NC* to *HC*, Bayesian ANCOVAs were performed on accuracy and reaction time in subtraction, including group (HC, PD-NC) and complexity (carry, non-carry) as fixed factors, and education, UPDRS-III, and depression as covariates (see Table 5B and Figure 1C). For *accuracy* (analysis B.I.a), the model including group, complexity, and participant was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity and participant, while there was inconclusive evidence regarding group and the interaction of group with complexity. For *reaction time* (analysis B.II.a), the model including group, complexity, participant, education and motor symptoms was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant, education, and group (HC being faster than PD-NC), moderate evidence for motor symptoms, and evidence against the interaction of group with complexity.

To compare *PD-MCI* to *PD-NC*, Bayesian ANCOVAs were performed on accuracy and reaction time in subtraction, including group (PD-NC, PD-MCI) and complexity (carry, non-carry) as fixed factors and LEDD and Hoehn & Yahr stage as covariates (see Table 5B and Figure 1D). For *accuracy* (analysis B.I.b), the model including group, complexity, and participant was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant and group (PD-NC being more accurate than PD-MCI) and evidence against including the interaction of group with complexity. For *reaction time* (analysis B.II.b), the model including group, complexity, participant, and Hoehn & Yahr disease stage was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant, group (PD-NC being faster than PD-MCI), and Hoehn & Yahr disease stage, while there was evidence against including the interaction of group with complexity.

1c) Multiplication

As a quality check, Bayesian paired *t*-tests were conducted in the HC group to test the complexity effect: There was evidence for accuracy differences ($BF_{10} > 1000$) with fewer errors for small (*M* = 0.99, *SD* = 0.01) than for large (*M* = 0.96, *SD* = 0.05) problems. There was also evidence for a reaction time difference ($BF_{10} > 1000$), with faster calculation for small (*M* = 1237 ms, *SD* = 432 ms) than for large (*M* = 2363 ms, *SD* = 1954 ms) problems. Additionally, ANCOVAs for both outcome measures indicated substantial evidence for the effect of complexity.

To compare *PD-NC* to *HC*, Bayesian ANOVAs were performed on accuracy and reaction time in multiplication, including group (HC, PD-NC) and complexity (carry, non-carry) as fixed factors and

education, UPDRS-III, and depression as covariates (see Table 5C and Figure 1E). For *accuracy* (analysis C.I.a), the model including complexity and motor symptoms was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, while there was evidence against group and the interaction of group with complexity. For *reaction time* (analysis C.II.a), the model including group, complexity, participant, education, and motor symptoms was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant, and motor symptoms, and moderate evidence for group (HC being faster than PD-NC) and education, while there was inconclusive evidence against the interaction of group with complexity.

To compare *PD-MCI to PD-NC*, Bayesian ANCOVAs were performed on accuracy and reaction time in multiplication, including group (PD-NC, PD-MCI) and complexity (carry, non-carry) as fixed factors and LEDD and Hoehn & Yahr stage as covariates (see Table 5C and Figure 1F). For *accuracy* (analysis C.I.b), the model including complexity was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, while there was evidence against group and the interaction of group with complexity. For *reaction time* (analysis C.II.b), the model including group, complexity, participant, and Hoehn & Yahr disease stage was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant, Hoehn & Yahr disease stage, and group (PD-NC being faster than PD-MCI), and evidence against the interaction of group with complexity.

1d) Division

As a quality check, Bayesian paired *t*-tests were conducted in the HC group to test the complexity effect: There was evidence for accuracy differences (BF_{10} = 139.59) with less errors for small (*M* = 0.99, *SD* = 0.02) than for large (*M* = 0.95, *SD* = 0.05) problems. There was also evidence for a reaction time difference (BF_{10} > 1000), with faster calculation for small (*M* = 1658 ms, *SD* = 747 ms) than for large (*M* = 2582 ms, *SD* = 1898 ms) problems. Additionally, ANCOVAs for both outcome measures indicated substantial evidence for the effect of complexity.

To compare *PD-NC* to *HC*, Bayesian ANCOVAs were performed on accuracy and reaction time in division, including group (HC, PD-NC) and complexity (carry, non-carry) as fixed factors and education, UPDRS-III, and depression as covariates (see Table 5D and Figure 1G). For *accuracy* (analysis D.I.a), the model including complexity, participant, and motor symptoms was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, and moderate evidence for motor symptoms, while there was evidence against group and the interaction of group with complexity. For *reaction time* (analysis D.II.a), the model including group,

complexity, participant, education, and motor symptoms was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant, education, motor symptoms, and group (HC being faster than PD-NC), while there was evidence against including the interaction of group with complexity.

To compare *PD-MCI* to *PD-NC*, Bayesian ANCOVAs were performed on accuracy and reaction time in division, including group (PD-NC, PD-MCI) and complexity (carry, non-carry) as fixed factors and LEDD and Hoehn & Yahr stage as covariates (see Table 5D and Figure 1H). For *accuracy* (analysis D.I.b), the model including complexity and Hoehn & Yahr stage was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity and moderate evidence for Hoehn & Yahr disease stage, while there was evidence against group and the interaction of group with complexity. For *reaction time* (analysis D.II.b), the model including group, complexity, participant, and Hoehn & Yahr disease stage was the most probable to explain the data compared to the null model. The analysis of effects yielded strong evidence for complexity, participant, Hoehn & Yahr disease stage, and group (PD-NC being faster than PD-MCI), and evidence against the interaction of group with complexity.

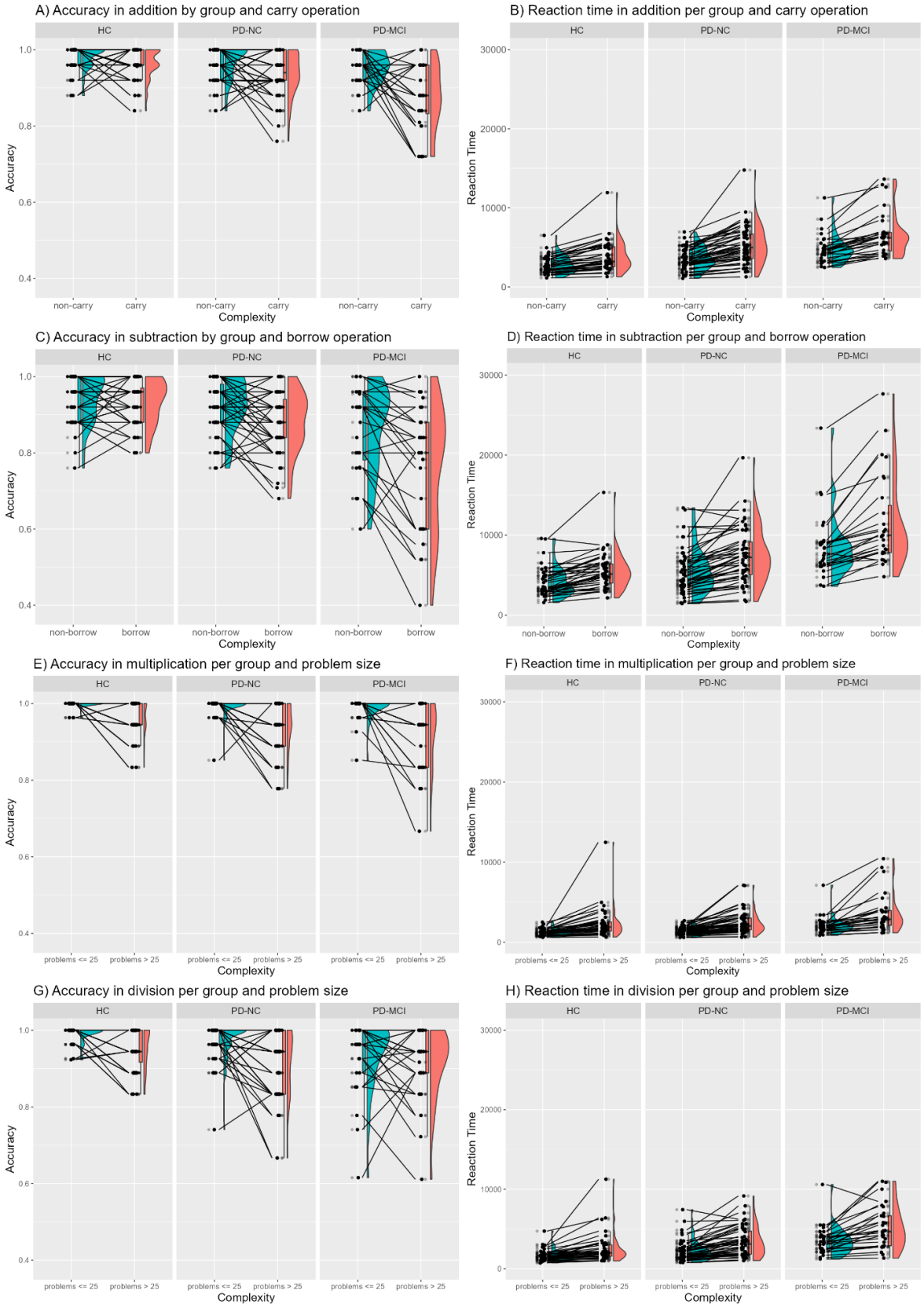

Figure 1. Distributions of accuracy and reaction time per group, task, and complexity condition.

Exploratory analysis on group differences in error categories

Two independent raters, with no involvement in the experimental procedures or statistical analyses of the data, classified errors according to established arithmetic error categories (as described in, e.g., Benavides-Varela et al., 2017; Roşca, 2009). Bayesian hypothesis testing yielded evidence against group differences in the occurrence of distinct error categories ($BF_{10} = 0.00$ for all comparisons, see Table 6). The most common error types for each arithmetic operation fell into the following subcategories: in addition, these comprised small unit error (28.3%), carry forgotten (21.3%), carry unnecessary (12.7%), carry twice (11.8%), and number persistence (11.8%). In subtraction, the most frequent were small unit error (22.3%), borrow forgotten (20.2%), borrow unnecessary (15.1%), and number persistence (11.1%). In multiplication, fact close (63.9%) and calculation error (21.3%) were most common. In division, fact close (35.2%), number persistence (23.9%), and rule for 1 (20.2%) were most prevalent. For explanations of the error categories, see the codebook on PsychArchive.

Table 6.

Frequencies of error categories per group and arithmetic operation.

Error category	HC	PD-NC	PD-MCI	Explanation & example
Addition	*n* = 58	*n* = 112	*n* = 144	
Procedural errors	46.6%	51.8%	41.0%	Wrong application of carry operation
Carry forgotten	15.5%	24.1%	21.5%	e.g., 67 + 26 = 83 (correct: 93)
Carry twice	12.1%	17.0%	7.6%	e.g., 48 + 16 = 74 (correct: 64)
Carry unnecessary	19.0%	10.7%	11.8%	e.g., 75 + 14 = 99 (correct: 89)
Formal errors	22.4%	12.5%	13.9%	Confusions in numbers and operations
Wrong operation	5.2%	1.8%	2.1%	e.g., 29 + 38 = 61 (correct: 67)
Number persistence	17.2%	10.7%	10.4%	e.g., 28 + 56 = 86 (correct: 84)
No decade	0.0%	0.0%	1.4%	e.g., 49 + 34 = 3 (correct: 83)
Calculation errors	31.0%	35.7%	45.1%	Small and large deviation from the result
Small unit error	22.4%	25.0%	33.3%	e.g., 17 + 45 = 63 (correct: 62)
Small decade error	6.9%	5.4%	4.9%	e.g., 41 + 26 = 57 (correct: 67)
Large error or unit & decade error	1.7%	5.4%	6.9%	e.g., 43 + 52 = 75 (correct: 95)
Subtraction	*n* = 126	*n* = 226	*n* = 236	
Procedural errors	46.0%	46.0%	39.4%	Wrong application of borrow operation
Borrow forgotten	16.7%	20.8%	21.6%	e.g., 51 – 37 = 24 (correct: 14)
Borrow twice	8.7%	10.2%	5.5%	e.g., 84 – 56 = 18 (correct: 28)
Borrow unnecessary	20.6%	15.0%	12.3%	e.g., 47 – 15 = 22 (correct: 32)
Formal errors	18.3%	20.8%	19.5%	Confusions in numbers and operations
Wrong operation	4.8%	4.0%	3.0%	e.g., 96 – 81 = 17 (correct: 15)
Inversion	4.0%	3.5%	4.2%	e.g., 41 – 27 = 26 (correct: 14)
Number persistence	9.5%	12.4%	10.6%	e.g., 63 – 16 = 46 (correct: 47)
No decade	0.0%	0.9%	1.7%	e.g., 93 – 36 = 7 (correct: 57)
Calculation errors	35.7%	33.2%	41.1%	Small and large deviation from the result
Small unit error	20.6%	23.0%	22.5%	e.g., 92 – 25 = 66 (correct: 67)
Small decade error	7.1%	1.8%	5.1%	e.g., 45 – 32 = 23 (correct: 13)
Large error or unit & decade error	7.9%	8.4%	13.6%	e.g., 47 – 15 = 52 (correct: 32)
Multiplication	*n* = 29	*n* = 65	*n* = 61	
Fact errors	82.8%	60.0%	70.5%	Errors within multiplication table
Fact close	75.9%	60.0%	62.3%	e.g., 3 × 9 = 24 (correct: 27)
Fact far	6.9%	0.0%	8.2%	e.g., 5 × 6 = 12 (correct: 30)
Formal errors	3.4%	15.4%	8.2%	Confusions in rules, numbers, operations
Rule for 1	0.0%	9.2%	6.6%	e.g., 3 × 1 = 1 (correct: 3)
Inversion	3.4%	1.5%	0.0%	e.g., 8 × 7 = 65 (correct: 56)
Number persistence	0.0%	4.6%	1.6%	e.g., 2 × 2 = 2 (correct: 4)
Calculation errors	13.8%	24.6%	21.3%	e.g., 9 × 6 = 62 (correct: 54)
Division	*n* = 44	*n* = 109	*n* = 94	
Fact errors	50.0%	48.6%	40.4%	Errors within division table
Fact close	38.6%	40.4%	27.7%	e.g., 54 ÷ 6 = 8 (correct: 9)
Fact far	11.4%	17.8%	12.8%	e.g., 36 ÷ 6 = 18 (correct: 6)
Formal errors	36.4%	41.1%	52.1%	Confusions in rules, numbers, operations
Rule for 1	6.8%	15.6%	31.9%	e.g., 9 ÷ 1 = 1 (correct: 9)
Operation	0.0%	1.8%	0.0%	e.g., 6 ÷ 2 = 4 (correct: 3)
Number persistence	29.5%	24.8%	20.2%	e.g., 35 ÷ 7 = 7 (correct: 5)
Calculation errors	13.6%	10.1%	7.4%	e.g., 45 ÷ 5 = 7 (correct: 9)

Research question 2) What role do domain-general factors play in group differences in arithmetic?

In order to identify the domain-general variable most strongly associated with accuracy and reaction time for each of the four basic arithmetic operations, correlations were calculated (see Table 7). The variable with the highest correlation was then included as a covariate in the ANCOVA with the respective dependent variable. The complete Bayesian ANCOVA tables for research question Q2 can be found in Supplementary Material E, while results detailing the changes in evidence due to the inclusion of the cognitive covariates are reported in the main text. Exploratory analyses highlighting that the domain-general factor most strongly associated with each arithmetic task varied by group can be found in Supplementary Material F, Tables S13 to S17 and Figure S1.

Table 7.

Pearson correlations between arithmetic performance and cognitive domains

	Addition		Subtraction		Multiplication		Division	
	ACC	RT	ACC	RT	ACC	RT	ACC	RT
Executive functions	*r* = -.08 [-.26; .10] *n* = 113	*r* = .14 [-.04; .31] *n* = 113	*r* = -.07 [-.25; .12] *n* = 112	*r* = .10 [-.09; .28] *n* = 112	*r* = -.28* [-.43; -.10] *n* = 115	*r* =-.10 [-.09; .27] *n* = 115	*r* = -.12 [-.30; .06] *n* = 115	*r* = .12 [-.07; .29] *n* = 115
Verbal working memory	***r* = .38****** **[.21; .52]** ***n* = 114**	*r* = -.39**** [-.53; -.22] *n* = 114	*r* = .24 [.06; .40] *n* = 113	*r* = -.37**** [-.51; -.19] *n* = 113	*r* = .17 [-.02; .34] *n* = 115	***r* = -.33***** **[-.47; -.15]** ***n* = 115**	*r* = .27* [.09; .42] *n* = 115	*r* = -.29** [-.44; -.11] *n* = 115
Visuo-spatial working memory	*r* = .26* [.08; .42] *n* = 115	***r* = -.39****** **[-.53; -.22]** ***n* = 115**	*r* = .35**** [.17; .49] *n* = 114	*r* = -.36**** [-.51; -.19] *n* = 114	*r* = .18 [.00; .35] *n* = 116	*r* = -.14 [-.31; .05] *n* = 116	*r* = .19 [.00; .35] *n* = 116	*r* = -.24* [-.40; -.06] *n* = 116
Attention	*r* = -.12 [-.30; .06] *n* = 113	*r* = .33*** [.15; .48] *n* = 113	*r* = -.24 [-.40; -.05] *n* = 112	*r* = .31*** [.13; .47] *n* = 112	*r* = -.11 [-.29; .07] *n* = 115	*r* = .20 [.01; .36] *n* = 115	*r* = -.06 [-.24; .12] *n* = 115	*r* = .24 [.05; .40] *n* = 115
Verbal memory	*r* = .24 [.05; .40] *n* = 112	*r* = -.18 [-.35; .01] *n* = 112	*r* = .33*** [.15; .48] *n* = 111	*r* = -.17 [-.35; .01] *n* = 111	*r* = .15 [-.03; .32] *n* = 113	*r* = -.09 [-.26; .10] *n* = 113	*r* = .32*** [.14; .47] *n* = 113	*r* = -.24 [-.40; -.05] *n* = 113
Language/ verbal reasoning	*r* = .25* [.06; .41] *n* = 114	*r* = -.34 [-.49; -.17] *n* = 114	***r* = .44****** **[.28; .58]** ***n* = 113**	***r* = -.44****** **[-.57; -.27]** ***n* = 113**	***r* = .31***** **[.13; .46]** ***n* = 115**	*r* = -.25* [-.41; -.07] *n* = 115	***r* = .54****** **[.39; .65]** ***n* = 115**	***r* = -.47****** **[-.60; -.31]** ***n* = 115**
Visuo-spatial function	*r* = .22 [.03; .38] *n* = 115	*r* = -.28* [-.43; -.10] *n* = 115	*r* = .19 [.01; .36] *n* = 114	*r* = -.33*** [-.48; -.16] *n* = 114	*r* = .16 [-.03; .32] *n* = 116	*r* = -.32*** [-.47; -.15] *n* = 116	*r* = .35**** [.18; .50] *n* = 116	*r* = -.33*** [-.48; -.15] *n* = 116

Note. *n* = overall sample size referring to the number of participants across all three groups who solved the task with sufficient accuracy to ensure task understanding. Values in bold indicate the correlation of highest magnitude for the respective arithmetic task. * $BF_{10} > 3$, ** $BF_{10} > 10$, *** $BF_{10} > 30$, **** $BF_{10} > 100$.

2a) Addition

When adding verbal working memory to the addition accuracy models, evidence for verbal working memory was inconclusive (BF_{incl} = 0.99 for HC vs. PD-NC, BF_{incl} = 1.42 for PD-NC vs. PD-MCI), and the category of evidence did not change for any of the variables. When adding visuo-spatial working memory to the reaction time models, there was substantial evidence for visuo-spatial working memory (BF_{incl} = 42.89 for HC vs. PD-NC, BF_{incl} = 50.95 for PD-NC vs. PD-MCI), but the category of evidence did not change for any of the variables (see Table S9).

2b) Subtraction

When adding verbal reasoning/language to the subtraction accuracy models, there was evidence for verbal reasoning/language in the model for HC and PD-NC groups (BF_{incl} = 4.61), while evidence was inconclusive in the model for PD-NC and PD-MCI groups (BF_{incl} = 0.96); the category of evidence did not change for any of the variables. When adding verbal reasoning/language to the reaction time models, evidence for including verbal reasoning/language was inconclusive (BF_{incl} = 1.57 for HC vs. PD-NC, BF_{incl} = 2.27 for PD-NC vs. PD-MCI), and the category of evidence did not change for any of the variables (see Table S10).

2c) Multiplication

When adding verbal reasoning/language to the multiplication accuracy models, evidence for verbal reasoning/language was inconclusive (BF_{incl} = 0.68 for HC vs. PD-NC, BF_{incl} = 1.83 for PD-NC vs. PD-MCI), and the category of evidence did not change for any of the variables. When adding verbal working memory to the reaction time model for HC and PD-NC groups, evidence for including verbal working memory was inconclusive (BF_{incl} = 2.93), and the category of evidence did not change for any of the variables. In the case of the reaction time model for PD-NC and PD-MCI groups, there was evidence for verbal working memory (BF_{incl} = 24.37), and the evidence for the group effect initially observed for research question Q1 turned into inconclusive evidence (BF_{incl} = 0.95, see Table S11).

2d) Division

When adding verbal reasoning/language to the division accuracy models, evidence for verbal reasoning/language was only present in one robustness check in the model for HC and PD-NC groups (BF_{incl} = 3.00) and substantial in the model for PD-NC and PD-MCI groups (BF_{incl} = 18.50). The category of evidence remained unchanged for all variables except for Hoehn & Yahr disease stage, which turned inconclusive in the model for PD-NC and PD-MCI groups (BF_{incl} = 0.77). When adding verbal

reasoning/language to the reaction time models, evidence for verbal reasoning/language was inconclusive (BF_{incl} = 1.56 for HC vs. PD-NC, BF_{incl} = 1.93 for PD-NC vs. PD-MCI) and the category of evidence did not change for any of the variables (see Table S12).

Research question 3) Which arithmetic tasks can successfully discriminate between healthy controls, PD patients without cognitive impairment and PD patients with mild cognitive impairment?

H3a) Can arithmetic performance be used to discriminate between PD-NC and PD-MCI?

A Bayesian logistic regression discriminating between PD-NC and PD-MCI indicated evidence for addition accuracy, multiplication reaction time, division accuracy and levodopa equivalence dose, yielding a correct classification rate of 80.82% (see Table 8).

Table 8.

Bayesian logistic regression on group membership of PD-NC vs. PD-MCI groups.

			95% Credible interval	
	Estimate	Estimation Error	Lower	Upper
Intercept	-3.71	1.09	-6.12	-1.83
Addition RT	-0.62	1.13	-2.89	1.56
Addition ACC	**-1.83**	**0.63**	**-3.17**	**-0.72**
Subtraction RT	-0.88	1.41	-3.79	1.74
Subtraction ACC	-0.96	0.65	-2.30	0.24
Multiplication RT	**3.43**	**1.10**	**1.47**	**5.80**
Multiplication ACC	-0.25	0.47	-1.22	0.64
Division RT	1.08	0.85	-0.53	2.80
Division ACC	**1.72**	**0.64**	**0.57**	**3.09**
Levodopa equivalence dose	**2.12**	**0.74**	**0.82**	**3.74**
Hoehn & Yahr	0.83	0.97	-1.02	2.80

Notes. All variables were *z*-standardized and centred. Model parameters whose 95% credibility intervals did not include zero were interpreted as meaningful and marked in bold.

H3b) Can arithmetic performance be used to discriminate between HC and PD-NC, despite them showing comparable results regarding global cognition?

The Bayesian logistic regression conducted to discriminate between HC and PD-NC did not converge and exhibited strong autocorrelations between model parameters, as well as a high R-hat value. Because the Bulk Effective Sample Size and Tail Effective Sample Size were too low, the posterior means, medians, variances, and tail quantiles derived from this regression may be unreliable. Therefore, the results of this regression will not be interpreted.

Discussion

The current study aimed to investigate arithmetic performance in PD by comparing patients across different disease stages (those with and without mild cognitive impairment) to an elderly control group. Accuracy results indicate that, across all four arithmetic operations, arithmetic performance is not affected by neurodegeneration at the PD-NC stage (see Box 2). Reaction times, on the other hand, are slowed in PD-NC compared to HC for subtraction, multiplication, and division. Neurodegeneration is associated with PD-MCI modulated number cognition, impacting addition accuracy, as well as both accuracy and reaction time in subtraction, and reaction time in multiplication and division. Both accuracy and reaction time were consistently driven by their respective domain-specific factors (carry, borrow, problem size effect) and a combination of clinical attributes. Furthermore, some cases demonstrated a joint contribution of a domain-general factor to arithmetic performance. These results will be further accentuated in the following discussion.

Box 2. *Summary of results for research questions 1 and 2*

	Addition	**Subtraction**	**Multiplication**	**Division**
HC vs. PD-NC	**→ inconclusive evidence for association between disease-specific neurodegeneration and performance in addition**	**→ association between disease-specific neurodegeneration and <u>slowed performance</u> in subtraction**	**→ association between disease-specific neurodegeneration and <u>slowed performance</u> with preserved accuracy in multiplication**	**→ association between disease-specific neurodegeneration and <u>slowed performance</u> with preserved accuracy in division**
	Accuracy HC ■ PD-NC + complexity **→ carry effect alone drives performance** Reaction Time HC ■ PD-NC + complexity + participant (random) + education years + motor symptoms **→ carry effect + visuo-spatial working memory drive performance**	Accuracy HC ■ PD-NC + complexity + participant (random) **→ borrow effect + language drive performance** Reaction Time HC < PD-NC + complexity + participant (random) + education years + motor symptoms **→ borrow effect alone drives performance**	Accuracy HC = PD-NC + complexity **→ problem size effect alone drives performance** Reaction Time HC < PD-NC + complexity + participant (random) + education years + motor symptoms **→ problem size effect alone drives performance**	Accuracy HC = PD-NC + complexity + *motor symptoms* **→ problem size effect + language drive performance** Reaction Time HC < PD-NC + complexity + *motor symptoms* **→ problem size effect alone drives performance**
PD-NC vs. PD-MCI	**→ association of <u>deficits</u> in addition with disease progression**	**→ association of <u>deficits</u> and <u>slowed performance</u> in subtraction with disease progression**	**→ association of <u>slowed performance</u> with preserved accuracy in multiplication with disease progression**	**→ association of <u>slowed performance</u> with preserved accuracy in division with disease progression**
	Accuracy PD-NC > PD-MCI + complexity **→ carry effect alone drives performance** Reaction Time PD-NC ■ PD-MCI + complexity + participant (random) + Hoehn & Yahr disease stage **→ carry effect + visuo-spatial working memory drive performance**	Accuracy PD-NC > PD-MCI + complexity + participant (random) **→ borrow effect alone drives performance** Reaction Time PD-NC < PD-MCI + complexity + participant (random) + Hoehn & Yahr disease stage **→ borrow effect alone drives performance**	Accuracy PD-NC = PD-MCI + complexity **→ problem size effect alone drives performance** Reaction Time *PD-NC < PD-MCI* + complexity + participant (random) + Hoehn & Yahr disease stage **→ problem size effect + verbal working memory drive performance**	Accuracy PD-NC = PD-MCI + complexity + *Hoehn & Yahr disease stage* **→ problem size effect + language drive performance** Reaction Time *PD-NC < PD-MCI* + complexity + participant (random) + Hoehn & Yahr disease stage **→ problem size effect alone drives performance**

Notes. "■" = inconclusive result, "<" or ">" = evidence for group effect, "=" = absence of evidence. Factors written in italics change evidence category after adding the domain-general factor in question 2. Conclusions written in grey relate to research question 2.

Distinct functioning of the four basic arithmetic operations

According to Dehaene (Dehaene, 1992, 1993; Dehaene et al., 1993, 2003; Dehaene & Cohen, 1995), arithmetic performance requires translation between verbal, magnitude, and visual number form representations. Extending this model, the two-network framework proposed by Klein and colleagues (2016) postulates a neural magnitude network (for difficult and less familiar problems) and a fact retrieval network (for simple and more familiar problems), which are anatomically distinct yet functionally integrated. Thus, PD patients may experience impairments in translating between codes, difficulties in processing or accessing specific representations, or disruptions within one or both of these arithmetic neural networks.

The current study found that accuracy in PD-MCI was impaired in two-digit addition and subtraction, which typically rely on the magnitude network, but preserved in single-digit multiplication and division, which primarily rely on the fact retrieval network. This aligns with McCloskey's view, suggesting that distinct, segregated networks underlying each basic arithmetic operation can be selectively affected (Dagenbach & McCloskey, 1992; McCloskey et al., 1991).

Despite these specific impairments, the overall high level of arithmetic performance in both PD groups indicates that PD patients retain the ability to carry out complex calculations and retrieve arithmetic facts. In contrast to basic numerical cognition, where reaction time slowing has only been observed at the PD-MCI stage (Loenneker et al., 2024), slowed reaction times were already evident in PD-NC patients for subtraction, multiplication, and division in the current sample.

Effects of PD pathology on complex arithmetic

Involved in complex and less familiar multi-digit calculations, the neuronal magnitude network includes the bilateral visual number form area, the intraparietal sulcus, supplementary motor areas, and the inferior frontal gyrus, with processing occurring via dorsal and ventral pathways (Klein et al., 2016; Klein & Knops, 2023). This network subserves two-digit arithmetic, which requires *place* × *value integration* (identification, activation and computation of distinct places and corresponding values, Klein et al., 2016; Nuerk et al., 2015). Basal ganglia seem to mediate rule application, which is necessary for complex calculation (Teichmann et al., 2005). Additionally, basal ganglia function affects the working memory resources required for complex arithmetic in PD (Brown et al., 1997; Fournet et al., 2000). The basal ganglia might also play a role in coordinating multiple cognitive operations, since PD patients have been shown to be increasingly impaired as the number of addends to be integrated in a sequential addition task increases (Saban et al., 2024). Consequently, there is a neurocognitive basis for our results on differences in accuracy in two-digit addition and subtraction between PD-NC and PD-MCI, suggesting

that arithmetic deficits in PD are associated with disease progression, and that processing is already slowed when comparing HC and PD-NC.

Comparing the current study's findings on arithmetic deficits with existing literature demonstrates its significant contribution to systematizing available evidence. Burgio et al. (2022) found impairments exclusively in written three-digit arithmetic, but not in written single-digit arithmetic, in a sample of 30% PD-MCI and 70% PD-NC patients. Given that the present study used orally presented problems, working memory capacity might have been taxed more strongly, leading to visible impairments in two-digit problems even in PD-MCI patients. Zamarian et al. (2006) reported arithmetic deficits only in more advanced tasks in a PD-NC sample, consistent with the slowed reaction times in our PD-NC sample, but no impairments regarding accuracy. Further evidence comes from studies reporting impairments in a graded difficulty arithmetic test, orally presenting addition and subtraction problems to PD patients without dementia (Scarpina et al., 2017), as well as an impaired mental calculation span in PD-NC (Gabrieli et al., 1996; Tamura et al., 2003). This underlines the current finding of calculation deficits for orally presented problems that require high working memory capacity. Our study's results align with the interpretation of PD-MCI as a predecessor stage of dementia, characterized by deficits in addition and subtraction, as PDD patients have shown deficits in three of the basic arithmetic operations (addition, subtraction, and multiplication) in written arithmetic, and across all four operations in mental arithmetic (Kalbe, 1999b). Collectively, this body of evidence suggests that complex arithmetic becomes progressively more affected with the advancement of neurodegeneration.

Arithmetic procedures in the place × value system are preserved in PD. Specifically, the carry and borrow effects, along with the procedural errors that occur when dealing with these operations (the most frequent error type in addition and subtraction), did not differ between HC, PD-NC and PD-MCI. The finding that place × value error types were the most frequent is in line with Loenneker et al. (2021), who analysed two simple word problems in a financial context from the Clinical Dementia Rating requiring division or sequential addition operations in PD-NC and PD-MCI patients. The items in the present study were overall more demanding than the items in Loenneker et al. (2021) and systematically manipulated regarding difficulty (i.e., the requirement to carry or borrow), thereby allowing more reliable insights. Interestingly, place × value processing remains intact in old age (Avcil & Artemenko, 2025), and this preservation appears to generalize to the neurodegenerative process of PD, both without and with mild cognitive impairment.

According to the "levodopa-overdose hypothesis" (Cools, 2006; Vaillancourt et al., 2013), dopaminergic overdosing of mesocortical projections to the prefrontal cortex can impair cognitive flexibility and attention. In line with this hypothesis, performance in a serial subtraction task was shown to decrease directly after levodopa intake (Dagan et al., 2021). While the serial subtraction task is simpler

than the one used in the present study, it is important to consider that arithmetic performance may have been influenced by participants' medication status, due to altered prefrontal activation following levodopa intake. This effect, which may have introduced further heterogeneity in performance within the two PD groups, could not be controlled for in the current experimental procedure, as participants were allowed to take their medication according to their regular regimen. Moreover, the predictor LEDD – which never was a predictor of arithmetic performance in the current data – only reflects dopaminergic degeneration as inferred from general dopaminergic medication requirements, but does not capture the medication state during task performance. However, because the PD-MCI group had a higher LEDD than the PD-NC group, frontal regions involved in the magnitude network might be more impaired by the levodopa-overdose, thereby affecting complex calculations in PD-MCI.

To summarize, complex arithmetic consisting of two-digit addition and subtraction problems seems to be impaired in PD-MCI. This impairment can be linked to the neural magnitude network.

Effects of PD pathology on arithmetic fact retrieval

The arithmetic fact network, engaged in simple and more familiar single-digit calculations, involves the left-hemispheric angular gyrus, visual number form area, retrosplenial cortex, middle temporal gyrus, medial frontal cortex, inferior frontal gyrus, prefrontal cortex, thalamus, arcuate fasciculus, and temporo-pontine tract, processed via dorsal and ventral pathways (Klein et al., 2016; Klein & Knops, 2023). Single-digit multiplication and division rely on this arithmetic fact retrieval network. All groups showed similar accuracy, suggesting preserved arithmetic fact retrieval in PD. This absence of deficits in arithmetic fact retrieval is in line with Burgio et al. (2022) and Zamarian et al. (2006), who similarly reported no impairments in single-digit arithmetic. However, this preserved accuracy was accompanied by a progressive slowing in reaction times from HC to PD-NC and from PD-NC to PD-MCI when solving multiplication and division problems. The role of the basal ganglia for arithmetic functioning – as targets of PD-specific neurodegeneration – has been attributed to their gating function in arithmetic fact retrieval (Delazer et al., 2004). Arguing within the framework of the triple code model, degeneration of the basal ganglia could disrupt the circuit mediating language-based retrieval of simple arithmetic facts (Dehaene & Cohen, 1997), thereby sparing accuracy but adversely affecting reaction times. We conclude that the arithmetic fact network seems to be intact, but accessing it becomes more time-consuming.

In multiplication and division, operand-related fact errors were the most frequent error type in all three groups. As the frequency of this error type did not differ between groups, we found evidence hinting at comparable arithmetic fact network structures in HC, PD-NC and PD-MCI. Additionally, the occurrence of rule-based errors (when 7 ÷ 7 is solved as 0 or 7 instead of as 1) indicated that all three

groups struggled to apply arithmetic rules correctly. Sensitivity to interference has been identified as a source of interindividual differences in arithmetic fact retrieval when solving simple multiplication problems (De Visscher et al., 2018), which can be linked to PD-specific executive deficits. The two patient groups might rely on different compensatory mechanisms, such as resorting to less efficient procedural strategies when arithmetic facts are no longer available for retrieval, leading to preserved accuracy but prolonged reaction times. Therefore, it is important to integrate accuracy and reaction time results, as the present study provides evidence for performance differences across all four basic arithmetic operations. The current study design does not allow us to disentangle which of the above-mentioned subprocesses (i.e., gating function of basal ganglia, arithmetic fact retrieval, application of arithmetic rules, sensitivity to interference, compensation) contributed to arithmetic deficits in PD-MCI.

Studies on healthy aging indicate that central arithmetic processes remain intact across the lifespan, but peripheral processes like domain-general cognitive functions, response demands, processing speed or working memory among others are affected by aging (for a meta-analysis see Artemenko et al., 2025). Linking these findings on healthy aging to PD-specific neurodegeneration, arithmetic procedures and fact retrieval seem to be preserved, whereas deficits in PD-NC and PD-MCI emerge when supportive cognitive processes, such as the speed of accessing arithmetic facts, come into play.

Primary or secondary arithmetic impairments in PD-MCI?

Results from research question 2 enable us to differentiate whether processes of arithmetic deficits in PD are primary, secondary, or indistinguishable. Across all four basic arithmetic operations, main effects of cognitive covariates and arithmetic complexity were observed, indicating a joint contribution of domain-general and -specific functions to complex arithmetic and arithmetic fact retrieval in PD. The observation that visuospatial working memory, verbal working memory, and verbal reasoning/language demonstrate the strongest association with performance in these four basic arithmetic operations in PD is in line with evidence from healthy young adults (Artemenko et al., 2018, 2019; Imbo, Vandierendonck, & De Rammelaere, 2007; Imbo, Vandierendonck, & Vergauwe, 2007). The role working memory plays in calculation becomes clear when considering its subprocesses, which include maintaining operands and intermediate results, updating, planning, and sequencing, all of which are essential for successful arithmetic performance (DeStefano & LeFevre, 2004).

Consistent and large effects of arithmetic complexity were found across operations (carry and borrow effects, problem size effects). Importantly, these effects did not differ between the three groups, despite their performance differences. This pattern is interesting, given that individuals with worse performance typically exhibit larger complexity effects than those with higher performance (Artemenko et al., 2018, 2019). Therefore, both healthy elderly and PD patients of different cognitive statuses

similarly process arithmetic complexity in a similar manner. This implies that neurodegeneration in PD might not have affected specific arithmetic procedures (place × value computations in terms of carry and borrow operations) or the retrieval of arithmetic facts (strategy shifts between simple and complex problems). Instead, this suggests that Parkinson's disease may affect domain-general cognitive processes rather than the domain-specific arithmetic processes that underlie arithmetic skills. This interpretation aligns with the findings of Zamarian and colleagues (2006), who observed that early PD patients without dementia were impaired only in complex mental arithmetic and calculation span tasks with high working memory load, leading them to conclude that arithmetic deficits in PD are secondary in nature.

Addition accuracy, multiplication reaction time, division accuracy and LEDD discriminated between the PD-NC and PD-MCI groups, suggesting that global cognitive and certain arithmetic impairments may share a common pathomechanism. However, there is no evidence supporting the use of arithmetic as an early marker for the detection of PD to discriminate between HC and PD-NC patients. Furthermore, the fact that group differences in multiplication reaction time became inconclusive after adding the domain-general variable visuo-spatial working memory indicates that this domain-general variable might be a better predictor of performance than motor disease staging (measured with Hoehn & Yahr stage) and cognitive staging (categorized according to the MoCA criterion).

Notably, additional exploratory analyses indicated that the domain-general factor most strongly associated with each arithmetic task differed depending on the group. In the case of verbal memory, associations with reaction time even diverged: PD-MCI patients with high verbal memory performance exhibited longer reaction times in addition, subtraction, and multiplication (i.e., positive correlation), whereas the PD-NC group exhibited shorter reaction times with high verbal memory performance (i.e., negative correlation). This disparity may relate to underlying compensatory mechanisms. PD-MCI patients with good verbal memory might have recruited these resources extensively, leading to longer processing times for arithmetic problems while maintaining high accuracy. By contrast, fast reaction times in the PD-NC group likely reflect fast and efficient access to stored arithmetic facts. If this interpretation holds, reaction times in arithmetic tasks might need to be interpreted differentially for each patient group.

We assume that domain-general cognitive functions, domain-specific arithmetic effects, and PD disease progression jointly influence arithmetic performance. Contrary to basic numerical cognition, where magnitude impairments seem to be primary (Loenneker et al., 2024), there is no evidence for specific arithmetic deficits independent of domain-general contributions.

Influence of further clinical markers of PD progression

When operationalizing PD progression using the Hoehn & Yahr disease stages, disease progression was also a predictor of addition reaction time and division accuracy differences between PD-NC and PD-MCI in cases where group itself was not a predictor. This indicates the need to differentiate between categorizations based on global cognitive performance and those based on motor deterioration. Additionally, Hoehn & Yahr disease stage explains reaction time differences in PD-NC and PD-MCI patients, while motor symptoms explain reaction time differences between HC and PD-NC across all arithmetic operations. Therefore, beyond the operationalization of disease progression based on cognitive staging, the progression marker Hoehn & Yahr stage shows the effect of disease progression on arithmetic performance in PD (comparable to basic numerical cognition in PD Loenneker et al., 2024). However, these effects might be, at least in part, attributable to the cognitive diagnosis, as more advanced disease stages are generally associated with cognitive deterioration in Parkinson's disease (Aarsland et al., 2021).

Lastly, lateralization of PD pathology might affect impairments at the behavioural level, since some arithmetic functions are also lateralized. For instance, arithmetic fact retrieval involves a predominantly left-lateralized neural network, while magnitude-related complex calculations rely on a bilateral network (e.g., Benavides-Varela et al., 2017; De Smedt et al., 2009; Klein & Knops, 2023). Therefore, a more fine-grained analysis of the respective clinical subtypes (right- or left-side onset) combined with neuroimaging data is necessary to fully understand arithmetic deficits in PD.

Limitations and perspectives

One limitation of the current study is the within- and between-group heterogeneity in the two patient groups. All RT models included the random factor participant, whereas none of the accuracy models did. This is especially interesting for the models without group effects, as this substantial interindividual variability in arithmetic skill might have masked group effects.

Next, most neuropsychological models of numerical cognition in clinical samples stem from lesion studies (e.g., Dehaene, 1992; Dehaene & Cohen, 1995b; Klein & Knops, 2023; McCloskey et al., 1991; Sokol et al., 1991) where the locus of impaired function can be roughly isolated and the onset of impairment is clearly defined. Contrary to these theoretical considerations, PD is a multi-system disorder in which neurotransmitter imbalance, amyloid plaque development, and brain atrophy jointly contribute to a distinct pattern of cognitive impairments, further compounded by PD-specific mood and motor disorders. Unlike focal lesions, PD is characterised by a gradual degeneration process rather than acute onset (Aarsland et al., 2021; Postuma et al., 2018). Therefore, generalizability from models of numerical cognition to the current PD sample is limited when trying to understand the underlying processes of arithmetic impairments.

Due to the heterogeneous nature of our sample, our maximum sample size of 119 might still have resulted in an underpowered study, even though it is 2-3 times larger than the samples in prior studies used for effect size estimation. However, the Bayes factor still allows for the interpretation of how likely the hypothesis is, given the data. As this study represents the first systematic study investigating arithmetic deficits in PD, our evidence can be used for sample size calculations in future studies focusing on a specific effect and conducted with an even larger sample.

Cognitive diagnosis of PD-NC and PD-MCI was made solely on the basis of MDS Task Force level I criteria, rather than a comprehensive level II neuropsychological test battery, as testing time was limited and priority was given to the target tasks of interest. This compromise may have increased the probability of misdiagnosing a patient's cognitive status, because the reliability and validity of level I criteria might be lower. To provide information on cognitive test performance, the number of patients scoring ≤ 1.5 standard values below the population mean, as specified in the respective test manuals, is reported in the Supplementary Material.

To reduce patient attrition, testing was split into two experimental sessions with sufficient breaks for recovery, and patients were allowed to take their medication during the sessions. Tests were conducted in a standardized order to mitigate potential order effects. However, this might have induced group differences if the three groups were differently affected by attrition, since advanced cognitive deficits were associated with longer experimental sessions with more breaks necessary for recovery. Although attritional and motivational effects are reduced by splitting the testing into two sessions, residual effects within a single session cannot be ruled out.

A further limitation of our study is its incomplete design, which includes only one matched healthy control group. Differences between PD-NC and PD-MCI do not necessarily relate to PD severity per se, but might also possibly result from additional neuropathological processes associated with alterations in patients' cognitive performance. Thus, arithmetic deficits in the PD-MCI group are difficult to interpret, as both general effects of cognitive impairments and/or specific disease progression potentially contribute to this effect. To clarify this issue, future studies should consider including an additional control group with non-PD-related MCI (neglected here due to limited resources). Moreover, it is unclear whether MCI and PD-MCI share the same pathomechanism, and therefore lead to the same deficit. Neurodegenerative cognitive impairments can be caused by brain atrophy, imbalances of cholinergic and dopaminergic neurotransmitters, amyloid pathology (typical of AD) and Lewy body pathology (typical of PD). Accordingly, MCI is associated with amyloid and cholinergic pathology, as in AD, whereas PD-MCI patients display a combination of amyloid, cholinergic, dopaminergic, and Lewy Body pathology (Chandra et al., 2019; C. H. Lin & Wu, 2015). Therefore, the interaction between PD

patient status and MCI may be due to the particular underlying pathomechanism, which would likely be unknown and/or heterogeneous for a HC-MCI group.

Given that Alzheimer's patients have demonstrated impairments in arithmetic fact retrieval, yet retain flexibility in strategy selection comparable to that of older adults (P. A. Allen et al., 2005; Duverne et al., 2003), it would be interesting to pursue this line of research in Parkinson's patients. Furthermore, deeper insights are needed to explore the efficiency of strategies employed. Patients and healthy controls may differ in their ability to flexibly apply the most suitable strategy, as efficient strategy switching declines across the lifespan (Hinault & Lemaire, 2016b; Uittenhove & Lemaire, 2015b) and might be linked to deteriorating executive functioning or a distinct reliance on working memory capacity (Fuson et al., 1997; Hinault & Lemaire, 2016b; Uittenhove & Lemaire, 2015b). Considering that Alzheimer's disease affects strategy execution and selection of optimal strategies only in the most difficult cases (Arnaud et al., 2008; Lemaire & Leclère, 2014), the effect of Parkinson's disease on arithmetic should be assessed in future research. Such assessments should employ paradigms such as a choice-no-choice design for retrieval and non-retrieval arithmetic strategies, or a rounding verification paradigm that incorporates a broad range of problem difficulties. It would also be interesting to add stages of subjective cognitive impairments as precursor stages of PD-MCI, as well as dementia as the final stage of cognitive decline in future investigations.

Conclusion

While arithmetic procedures and fact retrieval generally remain intact in Parkinson's disease, deficits were detected only in general cognitive processes subserving arithmetic once mild cognitive impairment was present. The deficits consisted of problems in two-digit addition and subtraction, as well as delayed arithmetic fact retrieval for single-digit multiplication and division. Interestingly, the deficits were not domain-specific, as arithmetic effects (problem size, carry, and borrow effects) did not differ between the groups. Given that the observed deficits occurred only at the PD-MCI stage, it is crucial for future research to classify PD patients based on their global cognition and to include individuals across the full cognitive continuum.

Acknowledgements

HL received stipends of the Landesgraduiertenfoerderung Baden-Wuerttemberg and of the Friedrich-Ebert Stiftung. HL was granted funding by the Wilhelm Schuler Foundation, which covered participant reimbursement and testing materials. CA was supported by the European Social Fund and the Ministry of Science, Research and the Arts of Baden-Wuerttemberg. CA and HCN are members of the LEAD Graduate School & Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Wuerttemberg. The authors would like to thank Krzysztof Cipora and Mojtaba Soltanlou for methodological and statistical advice, Morgan Hess and Nina Koehler for language proofreading, and Rahel Stahl, Lena Urbanzcyk, Martin Skopp, Katarina Sivec, Julia König, Annalena Wels, and Melina Ali Babaiyan Ghazvini for supporting participant testing. We would like to thank Yaren Findik for supporting the categorization of error types. The authors would like to thank Bruno Fimm for providing us with the TAP test.

For analyses in R, we used the libraries tidyverse (Wickham et al., 2019), gmodels (Warnes et al., 2018), table1 (Rich, 2023), car (Fox & Weisberg, 2019), lawstat (Gastwirth et al., 2023), Hmisc (Harrell Jr & Dupont, 2021), datawizard (Patil et al., 2022), rcompanion (Mangiafico, 2024), lattice (Sarkar, 2008), ggrain (M. Allen et al., 2021), psych (Revelle, 2021), GGally (Schloerke et al., 2024), BayesFactor (Morey et al., 2024), bayestestR (Makowski et al., 2024), afex (Singmann et al., 2024), cowplot (Wilke, 2024), data.table (Barrett et al., 2025) and ggpubr (Kassambara, 2025).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

All authors have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Conceptualization*: HDL, CA, HCN, and ILS; *Data Curation*, HDL; *Methodology*: HDL, CA, HCN, KW, and ILS; *Investigation*: HDL; *Formal Analysis*: HDL; *Writing – Original Draft*: HDL; *Writing – Review & Editing*: HDL, CA, HCN, KW, and ILS; *Project Administration*: HDL; *Visualization*: HDL; *Funding Acquisition*: HDL

References

Aarsland, D., Batzu, L., Halliday, G. M., Geurtsen, G. J., Ballard, C., Ray Chaudhuri, K., & Weintraub, D. (2021). Parkinson disease-associated cognitive impairment. *Nature Reviews Disease Primers 2021 7:1*, *7*(1), 1–21. https://doi.org/10.1038/s41572-021-00280-3

Aarsland, D., Creese, B., Politis, M., Chaudhuri, K. R., Ffytche, D. H., Weintraub, D., & Ballard, C. (2017). Cognitive decline in Parkinson disease. *Nature Reviews Neurology*, *13*(4), 217–231. https://doi.org/10.1038/nrneurol.2017.27

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Allen, P. A., Bucur, B., Lemaire, P., Duverne, S., Ogrocki, P. K., & Sanders, R. E. (2005). Influence of probable Alzheimer's disease on multiplication verification and production. *Aging, Neuropsychology, and Cognition*, *12*(1), 1–31. https://doi.org/10.1080/13825580490521322

Arcara, G., Burgio, F., Benavides-Varela, S., Toffano, R., Gindri, P., Tonini, E., Meneghello, F., & Semenza, C. (2019). Numerical Activities of Daily Living – Financial (NADL-F): A tool for the assessment of financial capacities. *Neuropsychological Rehabilitation*, *29*(7), 1062–1084. https://doi.org/10.1080/09602011.2017.1359188

Archambeau, K., De Visscher, A., Noël, M.-P., & Gevers, W. (2019). Impact of ageing on problem size and proactive interference in arithmetic facts solving. *Quarterly Journal of Experimental Psychology*, *72*(3), 446–456. https://doi.org/10.1177/1747021818759262

Ardila, A., & Rosselli, M. (2002). Acalculia and dyscalculia. *Neuropsychology Review*, *12*(4), 179–231. https://doi.org/10.1023/A:1021343508573

Arnaud, L., Lemaire, P., Allen, P., & Michel, B. F. (2008). Strategic aspects of young, healthy older adults', and Alzheimer patients' arithmetic performance. *Cortex*, *44*(2), 119–130. https://doi.org/10.1016/J.CORTEX.2006.03.001

Arsalidou, M., & Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009

Artemenko, C. (2021). Developmental fronto-parietal shift of brain activation during mental arithmetic across the lifespan: A registered report protocol. *PLoS ONE2*, *16*(8), e0256232. https://doi.org/10.1371/journal.pone.0256232

Artemenko, C., Avcil, M., Sautter, K., Lemaire, P., & Loenneker, H. D. (2025). Changes in arithmetic and numerical cognition during aging? A systematic review and meta-analysis. *Manuscript submitted for publication*.

Artemenko, C., Soltanlou, M., Bieck, S. M., Ehlis, A.-C., Dresler, T., & Nuerk, H.-C. (2019). Individual Differences in Math Ability Determine Neurocognitive Processing of Arithmetic Complexity: A Combined fNIRS-EEG Study. *Frontiers in Human Neuroscience*, *13*, 227. https://doi.org/10.3389/fnhum.2019.00227

Artemenko, C., Soltanlou, M., Dresler, T., Ehlis, A.-C., & Nuerk, H.-C. (2018). The neural correlates of arithmetic difficulty depend on mathematical ability: Evidence from combined fNIRS and ERP. *Brain Structure and Function*, *223*(6), 2561–2574. https://doi.org/10.1007/s00429-018-1618-0

Avcil, M., & Artemenko, C. (2025). Development of arithmetic across the lifespan: A registered report. *Developmental Psychology*, *61*(6), 1136–1151. https://doi.org/10.1037/dev0001566

Baayen, H. R., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Bangma, D. F., Tucha, O., Tucha, L., De Deyn, P. P., & Koerts, J. (2021). How well do people living with neurodegenerative diseases manage their finances? A meta-analysis and systematic review on the capacity to make financial decisions in people living with neurodegernative diseases. *Neuroscience & Biobehavioral Reviews*, *127*, 709–739. https://doi.org/10.1016/j.neubiorev.2021.05.021

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., Schwendinger, B., & Krylov, I. (2025). *data.table: Extension of „data.frame"* (Version R package version 1.17.99) [Software]. https://r-datatable.com

Benavides-Varela, S., Piva, D., Burgio, F., Passarini, L., Rolma, G., Meneghello, F., & Semenza, C. (2017). Re-assessing acalculia: Distinguishing spatial and purely arithmetical deficits in right-hemisphere damaged patients. *Cortex*, *88*, 151–164. https://doi.org/10.1016/j.cortex.2016.12.014

Benton, A. L., Varney, N. R., & Hamsher, K. deS. (1978). Visuospatial Judgment. *Archives of Neurology*, *35*(6), 364. https://doi.org/10.1001/archneur.1978.00500300038006

Bohnen, N. I., Kaufer, D. I., Ivanco, L. S., Lopresti, B., Koeppe, R. A., Davis, J. G., Mathis, C. A., Moore, R. Y., & DeKosky, S. T. (2003). Cortical Cholinergic Function Is More Severely Affected in Parkinsonian Dementia Than in Alzheimer Disease. *Archives of Neurology*, *60*(12), 1745. https://doi.org/10.1001/archneur.60.12.1745

Braak, H., Braak, E., Yilmazer, D., de Vos, R. A. I., Jansen, E. N. H., & Bohl, J. (1996). Pattern of brain destruction in Parkinson's and Alzheimer's diseases. *Journal of Neural Transmission*, *103*(4), 455–490. https://doi.org/10.1007/BF01276421

Braak, H., Del Tredici, K., Rüb, U., de Vos, R. A. I., Jansen Steur, E. N. H., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*, *24*(2), 197–211. https://doi.org/10.1016/S0197-4580(02)00065-9

Braak, H., Rüb, U., & Del Tredici, K. (2006). Cognitive decline correlates with neuropathological stage in Parkinson's disease. *Journal of the Neurological Sciences*, *248*(1–2), 255–258. https://doi.org/10.1016/J.JNS.2006.05.011

Braak, H., Rüb, U., Jansen Steur, E. N. H., Del Tredici, K., & De Vos, R. A. I. (2005). Cognitive status correlates with neuropathologic stage in Parkinson disease. *Neurology*, *64*(8), 1404–1410. https://doi.org/10.1212/01.WNL.0000158422.41380.82

Brown, L. L., Schneider, J. S., & Lidsky, T. I. (1997). Sensory and cognitive functions of the basal ganglia. *Current Opinion in Neurobiology*, *7*(2), 157–163. https://doi.org/10.1016/S0959-4388(97)80003-7

Burgio, F., Filippini, N., Weis, L., Danesin, L., Ferrazzi, G., Garon, M., Biundo, R., Facchini, S., Antonini, A., Benavides-Varela, S., Semenza, C., & Arcara, G. (2022). Neurocognitive correlates of numerical abilities in Parkinson's disease. *Neurological Sciences*, *43*(9), 5313–5322. https://doi.org/10.1007/s10072-022-06228-z

Caligiore, D., Helmich, R. C., Hallett, M., Moustafa, A. A., Timmermann, L., Toni, I., & Baldassarre, G. (2016). Parkinson's disease as a system-level disorder. *npj Parkinson's Disease*, *2*(1), 16025. https://doi.org/10.1038/npjparkd.2016.25

Cappelletti, M., Kopelman, M. D., Morton, J., & Butterworth, B. (2005). Dissociations in numerical abilities revealed by progressive cognitive decline in a patient with semantic dementia. *Cognitive Neuropsychology*, *22*(7), 771–793. https://doi.org/10.1080/02643290442000293

Chandra, A., Valkimadi, P.-E., Pagano, G., Cousins, O., Dervenoulas, G., Politis, M., & for the Alzheimer's Disease Neuroimaging Initiative. (2019). Applications of amyloid, tau, and neuroinflammation PET imaging to Alzheimer's disease and mild cognitive impairment. *Human Brain Mapping*, *40*, 5424–5442. https://doi.org/10.1002/hbm.24782

Chaudhuri, K. R., Martinez-Martin, P., Schapira, A. H. V., Stocchi, F., Sethi, K., Odin, P., Brown, R. G., Koller, W., Barone, P., MacPhee, G., Kelly, L., Rabey, M., MacMahon, D., Thomas, S., Ondo, W., Rye, D., Forbes, A., Tluk, S., Dhawan, V., … Olanow, C. W. (2006). International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for Parkinson's disease: The NMSQuest study. *Movement Disorders*, *21*(7), 916–923. https://doi.org/10.1002/mds.20844

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cohen-Mansfield, J., Skornick-Bouchbinder, M., & Brill, S. (2018). Trajectories of End of Life: A Systematic Review. *The Journals of Gerontology: Series B*, *73*(4), 564–572. https://doi.org/10.1093/geronb/gbx093

Collerton, D., Burn, D., McKeith, I., & O'Brien, J. (2003). Systematic Review and Meta-Analysis Show that Dementia with Lewy Bodies Is a Visual-Perceptual and Attentional-Executive Dementia. *Dementia and Geriatric Cognitive Disorders*, *16*(4), 229–237. https://doi.org/10.1159/000072807

Cools, R. (2006). Dopaminergic modulation of cognitive function-implications for l-DOPA treatment in Parkinson's disease. *Neuroscience & Biobehavioral Reviews*, *30*(1), 1–23. https://doi.org/10.1016/j.neubiorev.2005.03.024

Corsi, P. (1972). *Human memory and the medial temporal region of the brain*. McGill University (Canada).

Dablander, F., Bergh, D. van den, Ly, A., & Wagenmakers, E.-J. (2020). Default Bayes Factors for Testing the (In)equality of Several Population Variances. *arXiv*. http://arxiv.org/abs/2003.06278

Dagan, M., Herman, T., Bernad-Elazari, H., Gazit, E., Maidan, I., Giladi, N., Mirelman, A., Manor, B., & Hausdorff, J. M. (2021). Dopaminergic therapy and prefrontal activation during walking in individuals with Parkinson's disease: Does the levodopa overdose hypothesis extend to gait? *Journal of Neurology*, *268*(2), 658–668. https://doi.org/10.1007/s00415-020-10089-x

Dagenbach, D., & McCloskey, M. (1992). The organization of arithmetic facts in memory: Evidence from a brain-damaged patient. *Brain and Cognition*, *20*(2), 345–366. https://doi.org/10.1016/0278-2626(92)90026-I

de Lau, L. M. L., Verbaan, D., Marinus, J., & van Hilten, J. J. (2014). Survival in Parkinson's disease. Relation with motor and non-motor features. *Parkinsonism & Related Disorders*, *20*(6), 613–616. https://doi.org/10.1016/J.PARKRELDIS.2014.02.030

De Smedt, B., Grabner, R. H., & Studer, B. (2009). Oscillatory EEG correlates of arithmetic strategy use in addition and subtraction. *Experimental Brain Research*, *195*(4), 635–642. https://doi.org/10.1007/s00221-009-1839-9

De Visscher, A., & Noël, M.-P. (2014). Arithmetic facts storage deficit: The hypersensitivity-to-interference in memory hypothesis. *Developmental Science*, *17*(3), 434–442. https://doi.org/10.1111/desc.12135

De Visscher, A., Vogel, S. E., Reishofer, G., Hassler, E., Koschutnig, K., De Smedt, B., & Grabner, R. H. (2018). Interference and problem size effect in multiplication fact solving: Individual differences in brain activations and arithmetic performance. *NeuroImage*, *172*, 718–727. https://doi.org/10.1016/j.neuroimage.2018.01.060

Deary, I. J., Pattie, A., & Starr, J. M. (2013). The Stability of Intelligence From Age 11 to Age 90 Years: The Lothian Birth Cohort of 1921. *Psychological Science*, *24*(12), 2361–2368. https://doi.org/10.1177/0956797613486487

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (1993). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Hrsg.), *Sensorimotor Foundations of Higher Cognition* (S. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231447.003.0024

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Dehaene, S., & Cohen, L. (1995a). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Dehaene, S., & Cohen, L. (1995b). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Dehaene, S., & Cohen, L. (1997). Cerebral Pathways for Calculation: Double Dissociation between Rote Verbal and Quantitative Knowledge of Arithmetic. *Cortex*, *33*(2), 219–250. https://doi.org/10.1016/S0010-9452(08)70002-9

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Delazer, M., Domahs, F., Lochy, A., Karner, E., Benke, T., & Poewe, W. (2004). Number processing and basal ganglia dysfunction: A single case study. *Neuropsychologia*, *42*(8), 1050–1062. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2003.12.009

Delazer, M., Girelli, L., Granà, A., & Domahs, F. (2003). Number Processing and Calculation – Normative Data from Healthy Adults. *The Clinical Neuropsychologist*, *17*(3), 331–350. https://doi.org/10.1076/clin.17.3.331.18092

Delazer, M., Kemmler, G., & Benke, T. (2013). Health numeracy and cognitive decline in advanced age. *Aging, Neuropsychology, and Cognition*, *20*(6), 639–659. https://doi.org/10.1080/13825585.2012.750261

Delazer, M., Zamarian, L., Benke, T., Wagner, M., Gizewski, E. R., & Scherfler, C. (2019). Is an intact hippocampus necessary for answering 3 × 3? – Evidence from Alzheimer's disease. *Brain and Cognition*, *134*, 1–8. https://doi.org/10.1016/J.BANDC.2019.04.006

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*(3), 353–386. https://doi.org/10.1080/09541440244000328

DGN. (2016). *Leitlinien für Diagnostik und Therapie in der Neurologie. Idiopathisches Parkinson-Syndrom*. https://www.dgn.org/images/red_leitlinien/LL_2016/PDFs_Download/030010_LL_langfassung_ips_2016.pdf

Dirnberger, G., Frith, C. D., & Jahanshahi, M. (2005). Executive dysfunction in Parkinson's disease is associated with altered pallidal–frontal processing. *NeuroImage*, *25*(2), 588–599. https://doi.org/10.1016/J.NEUROIMAGE.2004.11.023

Domahs, F., Delazer, M., & Nuerk, H. (2006). What Makes Multiplication Facts Difficult. *Experimental Psychology*, *53*(4), 275–282. https://doi.org/10.1027/1618-3169.53.4.275

Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., & Nuerk, H. C. (2007). Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. *Behavioral and Brain Functions*, *3*(1), 66. https://doi.org/10.1186/1744-9081-3-66

Dorsey, E. R., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J. C., Ansha, M. G., Brayne, C., Choi, J.-Y. J., Collado-Mateo, D., Dahodwala, N., Do, H. P., Edessa, D., Endres, M., Fereshtehnejad, S.-M., Foreman, K. J., Gankpe, F. G., Gupta, R., Hankey, G. J., … Murray, C. J. L. (2018). Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *17*(11), 939–953. https://doi.org/10.1016/S1474-4422(18)30295-3

Duverne, S., Lemaire, P., & Michel, B. F. (2003). Alzheimer's disease disrupts arithmetic fact retrieval processes but not arithmetic strategy selection. *Brain and Cognition*, *52*(3), 302–318. https://doi.org/10.1016/S0278-2626(03)00168-4

Emre, M., Aarsland, D., Brown, R., Burn, D. J., Duyckaerts, C., Mizuno, Y., Broe, G. A., Cummings, J., Dickson, D. W., Gauthier, S., Goldman, J., Goetz, C., Korczyn, A., Lees, A., Levy, R., Litvan, I., McKeith, I., Olanow, W., Poewe, W., … Dubois, B. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders*, *22*(12), 1689–1707. https://doi.org/10.1002/mds.21507

Faulkenberry, T. (2019). *BF calculator for single-factor ANOVA summaries*.

Finke, M. S., Howe, J. S., & Huston, S. J. (2017). Old Age and the Decline in Financial Literacy. *Management Science*, *63*(1), 213–230. https://doi.org/10.1287/mnsc.2015.2293

Fournet, N., Moreaud, O., Roulin, J. L., Naegele, B., & Pellat, J. (2000). Working memory functioning in medicated Parkinson's disease patients and the effect of withdrawal of dopaminergic medication. *Neuropsychology*, *14*(2), 247–253. https://doi.org/10.1037/0894-4105.14.2.247

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression*. SAGE. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., Carpenter, T. P., & Fennema, E. (1997). Children's Conceptual Structures for Multidigit Numbers and Methods of

Multidigit Addition and Subtraction. *Journal for Research in Mathematics Education*, *28*(2), 130–162. https://doi.org/10.5951/jresematheduc.28.2.0130

Gabrieli, J. D. E., Singh, J., Stebbins, G. T., & Goetz, C. G. (1996). Reduced working memory span in Parkinson's disease: Evidence for the role of frontostriatal system in working and strategic memory. *Neuropsychology*, *10*(3), 322–332. https://doi.org/10.1037/0894-4105.10.3.321

Gastwirth, J. L., Gel, Y. R., Hui, W. L. W., Lyubchich, V., Miao, W., & Noguchi, K. (2023). *lawstat: Tools for Biostatistics, Public Policy, and Law*. https://doi.org/10.32614/CRAN.package.lawstat

Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211. https://doi.org/10.1080/19345747.2011.618213

Girelli, L., & Delazer, M. (2001). Numerical abilities in dementia. *Aphasiology*, *15*(7), 681–694. https://doi.org/10.1080/02687040143000122

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A. E., Lees, A., Leurgans, S., LeWitt, P. A., Nyenhuis, D., … LaPelle, N. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, *23*(15), 2129–2170. https://doi.org/10.1002/mds.22340

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia*, *47*(2), 604–608. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2008.10.013

Grice, J. W., Medellin, E., Jones, I., Horvath, S., McDaniel, H., O'lansen, C., & Baker, M. (2020). Persons as Effect Sizes. *Advances in Methods and Practices in Psychological Science*, *3*(4), 443–455. https://doi.org/10.1177/2515245920922982

Griffith, H. R., Stewart, C. C., Stoeckel, L. E., Okonkwo, O. C., Den Hollander, J. A., Martin, R. C., Belue, K., Copeland, J. N., Harrell, L. E., Brockington, J. C., Clark, D. G., & Marson, D. C. (2010). Magnetic resonance imaging volume of the angular gyri predicts financial skill deficits in people with amnestic mild cognitive impairment. *Journal of the American Geriatrics Society*, *58*(2), 265–274. https://doi.org/10.1111/j.1532-5415.2009.02679.x

Harrell Jr, F. E., & Dupont, C. (2021). *Hmisc: Harrell Miscellaneous* [Software]. https://cran.r-project.org/package=Hmisc

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). A metadata-driven methodology and workflow process for providing translational research informatics support. REDCap, Research electronic data capture. *J Biomed Inform.*, *42*(2), 377–381.

Hautzinger, M., Kühner, C., & Keller, F. (2006). *Das Beck Depressions Inventar II. Deutsche Bearbeitung und Handbuch zum BDI II*. Harcourt Test Services.

Helmstaedter, C., Lendt, M., & Lux, S. (2001). *VLMT - Verbaler Lern- und Merkfähigkeitstest*. Beltz.

Hinault, T., & Lemaire, P. (2016a). Age-related changes in strategic variations during arithmetic problem solving: The role of executive control. *Progress in Brain Research*, *227*, 257–276. https://doi.org/10.1016/BS.PBR.2016.03.009

Hinault, T., & Lemaire, P. (2016b). Age-related changes in strategic variations during arithmetic problem solving: The role of executive control. In M. Cappelletti & W. Fias (Hrsg.), *The Mathematical Brain Across the Lifespan* (S. 257–276). Elsevier.

Hirsch, L., Jette, N., Frolkis, A., Steeves, T., & Pringsheim, T. (2016). The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. *Neuroepidemiology*, *46*(4), 292–300. https://doi.org/10.1159/000445751

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

Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, *73*(21), 1738–1745. https://doi.org/10.1212/WNL.0b013e3181c34b47

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *Journal of neurology, neurosurgery, and psychiatry*, *55*(3), 181–184. https://doi.org/10.1136/JNNP.55.3.181

Hugo, J., & Ganguli, M. (2014). Dementia and cognitive impairment: Epidemiology, diagnosis, and treatment. *Clinics in geriatric medicine*, *30*(3), 421–442. https://doi.org/10.1016/j.cger.2014.04.001

Imbo, I., Vandierendonck, A., & De Rammelaere, S. (2007). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *Quarterly Journal of Experimental Psychology*, *60*(5), 708–731. https://doi.org/10.1080/17470210600762447

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychological Research*, *71*(4), 467–483. https://doi.org/10.1007/s00426-006-0044-8

Jackson, M., & Warrington, E. K. (1986). Arithmetic Skills in Patients with Unilateral Cerebral Lesions. *Cortex*, *22*(4), 611–620. https://doi.org/10.1016/S0010-9452(86)80020-X

Jankovic, J. (2008). Parkinson's disease: Clinical features and diagnosis. *Journal of neurology, neurosurgery, and psychiatry*, *79*(4), 368–376. https://doi.org/10.1136/jnnp.2007.131045

JASP Team. (2021). *JASP 0.16* (No. 0.9).

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press, Clarendon Press.

Jellinger, K. A. (2018). Dementia with Lewy bodies and Parkinson's disease-dementia: Current concepts and controversies. *Journal of Neural Transmission*, *125*(4), 615–650. https://doi.org/10.1007/s00702-017-1821-9

Jenkinson, C., Fitzpatrick, R., Peto, V., Greenhall, R., & Hyman, N. (1997). The Parkinson's Disease Questionnaire (PDQ-39): Development and validation of a Parkinson's disease summary index score. *Age and Ageing*, *26*(5), 353–357. https://doi.org/10.1093/ageing/26.5.353

Kalbe, E. (1999a). *Zahlenverarbeitung bei der Alzheimerschen Erkrankung und anderen Demenzen*. Universität Bielefeld.

Kalbe, E. (1999b). *Zahlenverarbeitung bei der Alzheimerschen Erkrankung und anderen Demenzen*. Universität Bielefeld.

Kalbe, E., & Kessler, J. (2002). Zahlenverarbeitungs- und Rechenstörungen bei Demenzen. *Zeitschrift für Gerontologie und Geriatrie*, *35*(2), 88–101. https://doi.org/10.1007/s003910200013

Kassambara, A. (2025). *ggpubr: „ggplot2" Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/authors.html

Kaufmann, L., Ischebeck, A., Weiss, E., Koppelstaetter, F., Siedentopf, C., Vogel, S. E., Gotwald, T., Marksteiner, J., & Wood, G. (2008). An fMRI study of the numerical Stroop task in individuals with and without minimal cognitive impairment. *Cortex*, *44*(9), 1248–1255. https://doi.org/10.1016/J.CORTEX.2007.11.009

Kazui, H., Kitagaki, H., & Mori, E. (2000). Cortical activation during retrieval of arithmetical facts and actual calculation: A functional magnetic resonance imaging study. *Psychiatry and Clinical Neurosciences*, *54*(4), 479–485. https://doi.org/10.1046/j.1440-1819.2000.00739.x

Kessels, R. P. C., Van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & De Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Klein, E., & Knops, A. (2023). The two-network framework of number processing: A step towards a better understanding of the neural origins of developmental dyscalculia. *Journal of Neural Transmission*, *130*(3), 253–268. https://doi.org/10.1007/s00702-022-02580-8

Klein, E., Suchan, J., Moeller, K., Karnath, H.-O., Knops, A., Wood, G., Nuerk, H.-C., & Willmes, K. (2016). Considering structural connectivity in the triple code model of numerical cognition: Differential connectivity for magnitude processing and arithmetic facts. *Brain Structure and Function*, *221*(2), 979–995. https://doi.org/10.1007/s00429-014-0951-1

Knops, A., Nuerk, H.-C., & Göbel, S. M. (2017). Domain-general factors influencing numerical and arithmetic processing. *Journal of Numerical Cognition*, *3*(2), 112–132. https://doi.org/10.5964/jnc.v3i2.159

Koob, A. O., Shaked, G. M., Bender, A., Bisquertt, A., Rockenstein, E., & Masliah, E. (2014). Neurogranin binds α-synuclein in the human superior temporal cortex and interaction is decreased in Parkinson's disease. *Brain research*, *1591*, 102–110. https://doi.org/10.1016/j.brainres.2014.10.013

Kowal, S. L., Dall, T. M., Chakrabarti, R., Storm, M. V., & Jain, A. (2013). The current and projected economic burden of Parkinson's disease in the United States. *Movement Disorders*, *28*(3), 311–318. https://doi.org/10.1002/mds.25292

Lemaire, P., & Leclère, M. (2014). Strategy selection in Alzheimer patients: A study in arithmetic. *Journal of Clinical and Experimental Neuropsychology*, *36*(5), 507–516. https://doi.org/10.1080/13803395.2014.911248

Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes*. Psychometrica. https://doi.org/10.13140/RG.2.2.17823.92329

Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the Couplings Among Intellectual Abilities and Constituent Cognitive Processes Across the Life Span. *Psychological Science*, *15*(3), 155–163. https://doi.org/10.1111/j.0956-7976.2004.01503003.x

Lin, C. H., & Wu, R. M. (2015). Biomarkers of cognitive decline in Parkinson's disease. *Parkinsonism & Related Disorders*, *21*(5), 431–443. https://doi.org/10.1016/J.PARKRELDIS.2015.02.010

Lin, S.-J., Baumeister, T. R., Garg, S., & McKeown, M. J. (2018). Cognitive Profiles and Hub Vulnerability in Parkinson's Disease. *Frontiers in Neurology*, *9*, 482. https://doi.org/10.3389/fneur.2018.00482

Lindenberger, U., & Baltes, P. B. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychology and Aging*, *12*(3), 410–432. https://doi.org/10.1037/0882-7974.12.3.410

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. https://doi.org/10.1177/0891988712466456

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

Loenneker, H. D., Artemenko, C., Willmes, K., Liepelt-Scarfone, I., & Nuerk, H.-C. (2024). Deficits in or Preservation of Basic Number Processing in Parkinson's Disease? A Registered Report. *Journal of Neuroscience Research*, *102*(11), e25397. https://doi.org/10.1002/jnr.25397

Loenneker, H. D., Becker, S., Nussbaum, S., Nuerk, H.-C., & Liepelt-Scarfone, I. (2021). Arithmetic Errors in Financial Contexts in Parkinson's Disease. *Frontiers in Psychology*, *12*, 1105. https://doi.org/10.3389/FPSYG.2021.629984

López González, I., Garcia-Esparcia, P., Llorens, F., & Ferrer, I. (2016). Genetic and Transcriptomic Profiles of Inflammation in Neurodegenerative Diseases: Alzheimer, Parkinson, Creutzfeldt-Jakob and Tauopathies. *International journal of molecular sciences*, *17*(2), 206. https://doi.org/10.3390/ijms17020206

Makowski, D., Lüdecke, D., Ben-Shachar, M. S., Patil, I., Wilson, M. D., Wiernik, B. M., Bürkner, P.-C., Mahr, T., Singmann, H., Gronau, Q. F., & Crawley, S. (2024). *bayestestR: Understand and Describe Bayesian Models and Posterior Distributions* (Version 0.14.0) [Software]. https://cran.r-project.org/web/packages/bayestestR/index.html

Mangiafico, S. S. (2024). *rcompanion: Functions to Support Extension Education Program Evaluation*. https://cran.r-project.org/web/packages/rcompanion/citation.html

Mantovan, M. C., Delazer, M., Ermani, M., & Denes, G. (1999). The Breakdown of Calculation Procedures in Alzheimer's Disease. *Cortex*, *35*(1), 21–38. https://doi.org/10.1016/S0010-9452(08)70783-4

Martin, R. C., Triebel, K. L., Kennedy, R. E., Nicholas, A. P., Watts, R. L., Stover, N. P., Brandon, M., & Marson, D. C. (2013). Impaired financial abilities in Parkinson's disease patients with mild cognitive impairment and dementia. *Parkinsonism & Related Disorders*, *19*(11), 986–990. https://doi.org/10.1016/J.PARKRELDIS.2013.06.017

Martinez-Horta, S., & Kulisevsky, J. (2019). Mild cognitive impairment in Parkinson's disease. *Journal of Neural Transmission*, *126*(7), 897–904. https://doi.org/10.1007/s00702-019-02003-1

Martini, L., Domahs, F., Benke, T., & Delazer, M. (2003). Everyday numerical abilities in Alzheimer's disease. *Journal of the International Neuropsychological Society*, *9*(6), 871–878. https://doi.org/10.1017/S1355617703960073

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McCloskey, M., Harley, W., & Sokol, S. M. (1991). Models of arithmetic fact retrieval: An evaluation in light of findings from normal and brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 377–397. https://doi.org/10.1037/0278-7393.17.3.377

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H. C. (2011). Early place-value understanding as a precursor for later arithmetic performance-A longitudinal study on numerical development. *Research in Developmental Disabilities*, *32*(5), 1837–1851. https://doi.org/10.1016/j.ridd.2011.03.012

Moeller, K., Willmes, K., & Klein, E. (2015). A review on functional and structural brain connectivity in numerical cognition. *Frontiers in Human Neuroscience*, *9*, 227. https://doi.org/10.3389/fnhum.2015.00227

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024). *BayesFactor: Computation of Bayes Factors for Common Designs*. https://doi.org/10.32614/CRAN.package.BayesFactor

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nuerk, H.-C., Moeller, K., & Willmes, K. (2015). Multi-digit Number Processing: Overview, Conceptual Clarifications, and Language Influences. In A. Kadosh, R., Dowker (Hrsg.), *The Oxford Handbook of Numerical Cognition* (S. 106–139). Oxford University Press.

Owen, A., Doyon, J., Dagher, A., Sadikot, A., & Evans, A. C. (1998). Abnormal basal ganglia outflow in Parkinson's disease identified with PET. Implications for higher cortical functions. *Brain*, *121*(5), 949–965. https://doi.org/10.1093/brain/121.5.949

Patil, I., Makowski, D., Ben-Shachar, M. S., Wiernik, B. M., Bacher, E., & Lüdecke, D. (2022). datawizard: An R Package for Easy Data Preparation andStatistical Transformations. *Journal of Open Source Software*, *7*(78), 4684. https://doi.org/10.21105/joss.04684

Petermann, F. (2012). *WAIS-IV. Wechsler Adult Intelligence Scale—Fourth Edition. Deutschsprachige Adaptation der WAIS-IV von D. Wechsler*. Pearson Assessment.

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., & Filos, S. (1982). Measurement of Functional Activities in Older Adults in the Community. *Journal of Gerontology*, *37*(3), 323–329. https://doi.org/10.1093/geronj/37.3.323

Pigott, K., Rick, J., Xie, S. X., Hurtig, H., Chen-Plotkin, A., Duda, J. E., Morley, J. F., Chahine, L. M., Dahodwala, N., Akhtar, R. S., Siderowf, A., Trojanowski, J. Q., & Weintraub, D. (2015).

Longitudinal study of normal cognition in Parkinson disease. *Neurology*, *85*(15), 1276–1282. https://doi.org/10.1212/WNL.0000000000002001

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., Halliday, G., Goetz, C. G., Gasser, T., Dubois, B., Chan, P., Bloem, B. R., Adler, C. H., & Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*, *30*(12), 1591–1601. https://doi.org/10.1002/mds.26424

Postuma, R. B., Poewe, W., Litvan, I., Lewis, S., Lang, A. E., Halliday, G., Goetz, C. G., Chan, P., Slow, E., Seppi, K., Schaffer, E., Rios-Romenets, S., Mi, T., Maetzler, C., Li, Y., Heim, B., Bledsoe, I. O., & Berg, D. (2018). Validation of the MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*, *33*(10), 1601–1608. https://doi.org/10.1002/mds.27362

R core team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research* [Software]. Northwestern University. https://cran.r-project.org/package=psych

Rich, B. (2023). *table1: Tables of Descriptive Statistics in HTML*. https://doi.org/10.32614/CRAN.package.table1

Rinne, J. O., Portin, R., Ruottinen, H., Nurmi, E., Bergman, J., Haaparanta, M., & Solin, O. (2000). Cognitive Impairment and the Brain Dopaminergic System in Parkinson Disease. *Archives of Neurology*, *57*(4), 470. https://doi.org/10.1001/archneur.57.4.470

Roşca, E. C. (2009). Arithmetic procedural knowledge: A cortico-subcortical circuit. *Brain Research*, *1302*, 148–156. https://doi.org/10.1016/j.brainres.2009.09.033

Saban, W., Pinheiro-Chagas, P., Borra, S., & Ivry, R. B. (2024). Distinct Contributions of the Cerebellum and Basal Ganglia to Arithmetic Procedures. *Journal of Neuroscience*, *44*(2). https://doi.org/10.1523/JNEUROSCI.1482-22.2023

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer. http://lmdvr.r-forge.r-project.org

Scarpina, F., Mauro, A., D'Aniello, G. E., Albani, G., Castelnuovo, G., Ambiel, E., & MacPherson, S. E. (2017). Cognitive Estimation in Non-demented Parkinson's Disease. *Archives of Clinical Neuropsychology*, *32*(4), 381–390. https://doi.org/10.1093/arclin/acx019

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Crowley, J. (2024). *GGally: Extension to „ggplot2"*. https://ggobi.github.io/ggally/

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin and Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Sherod, M. G., Griffith, H. R., Copeland, J., Belue, K., Krzywanski, S., Zamrini, E. Y., Harrell, L. E., Clark, D. G., Brockington, J. C., Powers, R. E., & Marson, D. C. (2009). Neurocognitive predictors of financial capacity across the dementia spectrum: Normal aging, mild cognitive impairment, and Alzheimer's disease. *Journal of the International Neuropsychological Society : JINS*, *15*(2), 258–267. https://doi.org/10.1017/S1355617709090365

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., Love, J., Lenth, R., & Christensen, R. H. B. (2024). *afex: Analysis of Factorial Experiments* (Version 1.3-1) [Software]. https://cran.r-project.org/web/packages/afex/index.html

Sokol, S. M., McCloskey, M., Cohen, N. J., & Aliminosa, D. (1991). Cognitive representations and processes in arithmetic: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 355–376. https://doi.org/10.1037/0278-7393.17.3.355

Stazyk, E. H., Ashcraft, M. H., & Hamann, M. S. (1982). A network approach to mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 320–335. https://doi.org/10.1037/0278-7393.8.4.320

Tamura, I., Kikuchi, S., Otsuki, M., Kitagawa, M., & Tashiro, K. (2003). Deficits of working memory during mental calculation in patients with Parkinson's disease. *Journal of the Neurological Sciences*, *209*(1–2), 19–23. https://doi.org/10.1016/S0022-510X(02)00457-4

Teichmann, M., Dupoux, E., Kouider, S., Brugières, P., Boissé, M.-F., Baudic, S., Cesaro, P., Peschanski, M., & Bacoud-Lévi, A.-C. (2005). The role of the striatum in rule application: The model of Huntington's disease at early stage. *Brain*, *128*(5), 1155–1167. https://doi.org/10.1093/brain/awh472

Thomann, A. E., Goettel, N., Monsch, R. J., Berres, M., Jan, T., Steiner, L. A., & Monsch, A. U. (2018). The Montreal Cognitive Assessment: Normative Data from a German Speaking Cohort and Comparison with International Normative Samples. *Journal of Alzheimer's Disease*, *64*, 643–655.

Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., & Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders*, *25*(15), 2649–2653. https://doi.org/10.1002/mds.23429

Tucker-Drob, E. M. (2019). Cognitive Aging and Dementia: A Life-Span Perspective. *Annual Review of Developmental Psychology*, *1*(1), 177–196. https://doi.org/10.1146/annurev-devpsych-121318-085204

Tucker-Drob, E. M., Brandmaier, A. M., & Lindenberger, U. (2019). Coupled cognitive changes in adulthood: A meta-analysis. *Psychological Bulletin*, *145*(3), 273–301. https://doi.org/10.1037/bul0000179

Uittenhove, K., & Lemaire, P. (2015a). The effects of aging on numerical cognition. In R. Cohen Kadosh & A. Dowker (Hrsg.), *The Oxford Handbook of Numerical Cognition* (S. 345–366). Oxford University Press.

Uittenhove, K., & Lemaire, P. (2015b). The effects of aging on numerical cognition. In R. Cohen Kadosh & A. Dowker (Hrsg.), *The Oxford Handbook of Numerical Cognition* (S. 345–366). Oxford University Press.

UNESCO. (2012). *International Standard Classification of Education ISCED 2011*.

Uribe, C., Segura, B., Baggio, H. C., Campabadal, A., Abos, A., Compta, Y., Marti, M. J., Valldeoriola, F., Bargallo, N., & Junque, C. (2018). Differential Progression of Regional Hippocampal Atrophy in Aging and Parkinson's Disease. *Frontiers in Aging Neuroscience*, *10*, 325. https://doi.org/10.3389/fnagi.2018.00325

Vaillancourt, D. E., Schonfeld, D., Kwak, Y., Bohnen, N. I., & Seidler, R. (2013). Dopamine overdose hypothesis: Evidence and clinical implications. *Movement Disorders*, *28*(14), 1920–1929. https://doi.org/10.1002/mds.25687

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2020). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, 1–14. https://doi.org/10.3758/s13423-020-01798-5

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory and Cognition*, *33*(1), 1–16. https://doi.org/10.3758/BF03195293

Warnes, G. R., Bolker, B., Lumley, T., & Johnson, R. C. (2018). *gmodels: Various R Programming Tools for Model Fitting* [Software]. https://cran.r-project.org/package=gmodels

Wechsler, D. (1995). *Wechsler Adult Intelligence Scale*. The Psychological Corporation.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilke, C. O. (2024). *cowplot: Streamlined Plot Theme and Plot Annotations for „ggplot2"* (Version 1.1.3) [Software]. https://cran.r-project.org/web/packages/cowplot/index.html

Willmes, K., Klein, E., & Nuerk, H.-C. (2013). Akalkulie. In *Funktionelle MRT in Psychiatrie und Neurologie* (S. 577–586). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29800-4_36

Wood, G., Ischebeck, A., Koppelstaetter, F., Gotwald, T., & Kaufmann, L. (2009). Developmental trajectories of magnitude processing and interference control: An fMRI study. *Cerebral Cortex*, *19*(11), 2755–2765. https://doi.org/10.1093/cercor/bhp056

Yang, Y., Zhong, N., Friston, K., Imamura, K., Lu, S., Li, M., Zhou, H., Wang, H., Li, K., & Hu, B. (2017). The functional architectures of addition and subtraction: Network discovery using fMRI and DCM. *Human Brain Mapping*, *38*(6), 3210–3225. https://doi.org/10.1002/hbm.23585

Zamarian, L., Visani, P., Delazer, M., Seppi, K., Mair, K. J., Diem, A., Poewe, W., & Benke, T. (2006). Parkinson's disease and arithmetics: The role of executive functions. *Journal of the Neurological Sciences*, *248*(1–2), 124–130. https://doi.org/10.1016/J.JNS.2006.05.037

Zbrodoff, N. J., & Logan, G. D. (2005). What Everyone Finds. The Problem-Size Effect. In *Handbook of mathematical cognition* (S. 331–346). Taylor & Francis Group.

Zhou, X., Li, M., Li, L., Zhang, Y., Cui, J., Liu, J., & Chen, C. (2018). The semantic system is involved in mathematical problem solving. *NeuroImage*, *166*, 360–370. https://doi.org/10.1016/J.NEUROIMAGE.2017.11.017

Zimmermann, P., & Fimm, V. (2017). *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* (No. 2.3.1). Psychologische Testsysteme.

Supplementary Material for

Arithmetic deficits in Parkinson's Disease?

– A Registered Report

Hannah D. Loenneker, Inga Liepelt-Scarfone, Klaus Willmes, Hans-Christoph Nuerk, & Christina Artemenko

Supplementary Material A: Deviations between stage 1 and stage 2

Piloting

The stage 1 of this registered report did not mention if and how the experimental procedure was piloted. Two healthy elderly piloted the entire testing protocol (for details see Supplementary Material of Loenneker et al., 2024). Regarding the current article, piloting did not lead to changes in the reported measures. We added the following specification to the section *Procedure*:

"Piloting the entire experimental procedure with two healthy elderly persons did not lead to any changes regarding the measures reported in the current article."

Sampling

In the stage 1 of this registered report, a sequential Bayes factor design with a maximum sample size of n_{max} = 120 valid data sets was established. Due to time constraints and patient availability, testing had to be stopped when 119 valid datasets were reached after a recruitment period of 17 months with Covid-19-related interruptions. The factor of interest group reached a value of $BF_{10} \geq 6$ in 7 analyses and a value of $BF_{10} \geq 3$ in one analysis (i.e., moderate to large evidence in favour of a group effect), a $BF_{10} \leq 1/3$ in 4 analyses (i.e., evidence for the absence of a group effect) and an inconclusive result in 4 analyses for Q1. For Q2, the factor group reached a value of $BF_{10} \geq 6$ in 3 analyses and a value of $BF_{10} \geq 3$ in 3 analyses (i.e., moderate to large evidence in favour of a group effect), a $BF_{10} \leq 1/3$ in 4 analyses (i.e., evidence for the absence of a group effect) and an inconclusive result in 6 analyses. Results of the respective robustness analyses indicate that evidence remains largely the same under a range of scale factors. Therefore, the result of 4 analyses in Q1 and 6 analyses in Q2 would have needed further recruitment to yield conclusive results. It is unlikely that a 120th participant would have been sufficient to reach this level of conclusiveness (see Bayesian ANCOVA tables in the main text).

Pre-processing

We originally wrote the following sentence in the stage 1 of this registered report:

"To match the scale of our prior distribution (Cauchy with a scale factor of 1 in Cohen's d units), we will hence transform both RT and ACC data into Cohen's d."

As the ANCOVA models we preregistered incorporate a group difference using a categorical factor of group, we cannot transform the dependent variable to a scale of Cohen's units, which would mean a standardized group difference. Therefore, this sentence was deleted from the manuscript and the dependent variables are measured in milliseconds in case of reaction time and percentage in case of accuracy.

It only occurred to us when preprocessing the actual data that the last steps of the preprocessing pipeline suggested by Baayen & Milin (2010) cannot be realized for our data. The reason is that we use aggregated data as needed for ANCOVA models and thus our analysis does not allow for including previous trial reaction time into the model. This would only be possible in generalized linear mixed models which are based on trial data of participants rather than aggregated data.

The manuscript text was changed as follows:

"**Reaction times.** Data trimming for RTs in the arithmetic tasks is adapted from Baayen & Milin (2010). The RT distribution of correctly solved trials were inspected with by-subject quantile-quantile plots, in order to identify the best suited theoretical model for data transformation to approach normally distributed data. The most accurate distribution was determined with the best model fit (see Suppplementary Material B, Table S1 for the respective distributions) and used to transform the data. After that, outliers were excluded in two steps. First, anticipations were excluded defined as RTs faster than 200 ms. Second, model criticism was used to exclude remaining outliers, based on Shapiro tests for normality. Those data points with absolute standardized residuals exceeding 3 were removed. For participants to be included in the RT analysis, a minimum of five valid data points out of 20-25 trials needed to be available per condition (i.e., minimum ACC of 20-25%). It was further preregistered that temporal dependencies should be corrected for with autocorrelation functions by participant and a regression model fitted to responses with a log-transformation for latencies and including the covariates trial number and preceding RT. However, this step could not be conducted, as it is incompatible with hypothesis testing using ANCOVAs which are running on averaged data (only appropriate for generalized linear models considering each observation of each subject; see Supplementary Material A). An overview of chosen transformations and excluded observations per preprocessing step can be found in the Supplementary Material B (Table S1)."

Wording in the material section

We became aware of the fact that our description of the "Functional Activities Questionnaire" was not precise enough, as we initially stated that it assesses social function. A more appropriate description in this case is "activities of daily living". The manuscript text was changed as follows:

In Table 2, "social function" was replaced by "activities of daily living".

"The patients' current *activities of daily living* were measured using the Functional Activities Questionnaire (FAQ; Pfeffer et al., 1982), which is tailored to older adults. Participants must rate their level of performance (ranging from 0 = normal to 3 = dependent) in 10 different daily life activities, resulting in a sum score characterising ADL.

Health-related quality of life was assessed with the single index score of the 39-item Parkinson's Disease Questionnaire (PDQ-39; Jenkinson et al., 1997). The eight dimensions such as activities of daily living and social support were coded on a scale ranging from 0 (= perfect health) to 100 (= worst health).

Caregiver assessment. For a broader evaluation of the participant's ADL, the FAQ was also completed by a caregiver. […]"

All of these changes are incorporated in the stage II registered report.

Reporting of results

Originally, we wrote the following: "BF_{10} indicated the probability of evidence in favour of the alternative hypothesis and BF_{01} indicated the probability of evidence in favour of the null hypothesis with $BF_{01} = 1/BF_{10}$ (which will be additionally reported when $BF_{10} < 1$)." As the current article reports a lot of different results, we stepped back from additionally reporting BF_{01} for results with evidence against a group effect. Instead, for better readability, we marked all those values in italics in the respective tables. Therefore, the part" (which will be additionally reported when $BF_{10} < 1$)" has been deleted.

We originally preregistered: "After data acquisition, robustness of the BF across different scale factors will be assessed with a robustness plot in JASP." However, JASP currently does not allow for robustness analyses in ANCOVAs, which is why we report the robustness checks in tables instead of plotting them. Therefore, the sentence was changed as follows: "After data acquisition, robustness of the BF across different scale factors was assessed and can be found in the respective ANCOVA tables."

We now refer to the published stage 2 version of the other registered report in this research project instead of to stage 1:

Loenneker, H. D., Artemenko, C., Willmes, K., Liepelt-Scarfone, I., & Nuerk, H.-C. (2024). Deficits in or Preservation of Basic Number Processing in Parkinson's Disease? A Registered Report. *Journal of Neuroscience Research*, 102(11), e25397. https://doi.org/10.1002/jnr.25397

Supplementary Material B: Details on preprocessing

Table S1.

Exclusions and data transformations prior to data analysis.

Task	Dependent variable	Participants not completing the task	Participants excluded exceeding 3 *SD* below group mean accuracy	Remaining % of trials*	Violated assumptions in raw data	Data transformation
Addition	Accuracy	2 PD-NC	1 HC, 1 PD-MCI	98.25%	Normal distribution, homogeneity of variances	Arcsine
	Reaction Time			98.05%	Normal distribution	Log
Subtraction	Accuracy	3 PD-NC, 2 PD-MCI		99.91%	Normal distribution, homogeneity of variances	Arcsine
	Reaction Time			98.15%	Normal distribution, homogeneity of variances	Log
Multiplication	Accuracy	1 HC, 1 PD-NC	1 PD-MCI	99.11%	Normal distribution, homogeneity of variances	Logit
	Reaction Time			97.72%	Normal distribution	Log
Division	Accuracy	1 HC, 2 PD-NC		99.94%	Normal distribution, homogeneity of variances	Logit
	Reaction Time			97.56%	Normal distribution, homogeneity of variances	Log

Note. *The remaining % of trials for accuracy is the proportion of participants excluded exceeding 3 SD below group mean accuracy divided by all participants. The remaining % of trials for reaction time is the proportion after excluding trials based on: the difference between keypress and key release exceeds 3 *SD* around the individual mean, anticipations < 200 ms, and trimming observations exceeding 3 *SD* around the individual mean of participants without normally distributed data after transformation; divided by all correct trials of participants not exceeding 3 *SD* below accuracy group mean.

Supplementary Material C: Further group characteristics

Table S2.

Descriptive experimental data per group.

	HC	PD-NC	PD-MCI	Cohen's d	
Addition	*n* = 39	*n* = 48	*n* = 28	HC – PD-NC	PD-NC – PD-MCI
Accuracy (%)					
Mean (*SD*)	97.0 (3.0)	95.0 (4.0)	92.0 (4.0)	.11	.11
Median (*MAD*), Range	98.0 (3.0), 90.0 – 100	96.0 (3.0), 86.0 – 100	92.0 (3.0), 82.0 – 98.0		
Skewness, Kurtosis	-0.88, 0.25	-0.51, -0.83	-0.37, -0.68		
Reaction Time (ms)					
Mean (*SD*)	3362 (1504)	4435 (1844)	5675 (2234)	-.11	-.10
Median (*MAD*), Range	3005 (1340), 1209 – 9277	4197 (1858), 1170 – 10501	5301 (1630), 3020 – 11906		
Skewness, Kurtosis	1.56, 3.8	0.66, 0.72	1.23, 0.91		
Subtraction	*n* = 40	*n* = 47	*n* = 27		
Accuracy (%)					
Mean (*SD*)	94.0 (5.0)	90.0 (6.0)	81.0 (13.0)	.13	.22
Median (*MAD*), Range	94.0 (6.0), 80.0 – 100	92.0 (6.0), 76.0 – 100	86.0 (15.0), 54.0 – 98.0		
Skewness, Kurtosis	-0.7, 0.29	-0.57, -0.67	-0.51, -0.96		
Reaction Time (ms)					
Mean (*SD*)	4837 (1989)	6522 (3032)	10046 (4781)	-.13	-.17
Median (*MAD*), Range	4306 (1776), 1866 – 12324	5944 (2489), 1572 – 14091	8394 (2532), 4145 – 25659		
Skewness, Kurtosis	1.47, 3.16	0.61, -0.29	1.5, 2.09		
Multiplication	*n* = 39	*n* = 49	*n* = 28		
Accuracy (%)					
Mean (*SD*)	98.0 (2.0)	97.0 (3.0)	96.0 (4.0)	.08	.05
Median (*MAD*), Range	98.0 (3.0), 93.0 – 100	98.0 (3.0), 89.0 – 100	98.0 (3.0), 84.0 – 100		
Skewness, Kurtosis	-0.79, -0.3	-0.91, 0.09	-1.19, 0.56		
Reaction Time (ms)					
Mean (*SD*)	1663 (923)	1812 (771)	2673 (1519)	-.03	-.16
Median (*MAD*), Range	1431 (571), 632 – 5833	1714 (644), 629 – 4021	2373 (810), 989 – 8430		
Skewness, Kurtosis	2.5, 8.29	0.95, 0.47	2.07, 5.02		
Division	*n* = 39	*n* = 48	*n* = 29		
Accuracy (%)					
Mean (*SD*)	97.0 (3.0)	95.0 (5.0)	92.0 (8.0)	.11	.09
Median (*MAD*), Range	98.0 (3.0), 91.0 – 100	96.0 (3.0), 78.0 – 100	96.0 (7.0), 71.0 – 100		
Skewness, Kurtosis	-0.75, -0.84	-1.16, 1.25	-1.25, 0.8		
Reaction Time (ms)					
Mean (*SD*)	2015 (1125)	2715 (1469)	4153 (1941)	-.14	-.30
Median (*MAD*), Range	1727 (554), 816 – 6211	2313 (1137), 859 – 6955	4029 (1683), 1255 – 9562		
Skewness, Kurtosis	2.1, 4.45	1.13, 0.62	0.82, 0.23		

Note. Descriptive statistics were run on the preprocessed data used for inferential statistics as reported in the main text.

Table S3.

Cognitive characteristics per group.

	HC	PD-NC	HC vs. PD-NC BF_{10}	PD-MCI	PD-NC vs. PD-MCI BF_{10}
MoCA	27.4 (1.03)	27.8 (1.16)	0.68	22.9 (1.93)	**> 1000**
Executive functions					
Raw score (errors)	0.44 (0.85)	1.66 (2.88)	**3.77**	1.64 (2.09)	0.24
z-score	-0.30 (0.32)	-0.65 (0.71)	**7.74**	-0.69 (0.68)	0.25
Verbal working memory					
Raw score (span)	4.35 (1.03)	4.26 (1.08)	0.24	3.07 (1.39)	**290.88**
Visuo-spatial working memory					
Raw score (span)	5.60 (0.67)	5.26 (1.08)	0.82	5.00 (0.71)	0.43
z-score	0.53 (0.61)	0.26 (0.98)	0.63	0.03 (0.62)	0.41
Attention					
Raw score (msec)	314 (115)	351 (95.4)	0.72	386 (117)	0.59
z-score	-0.80 (0.93)	-1.30 (0.85)	**5.12**	-1.68 (0.62)	1.39
Verbal memory					
Raw score (sum score)	9.95 (3.30)	9.10 (3.12)	0.44	6.12 (3.01)	**158.16**
z-score	-0.32 (1.08)	-0.58 (1.06)	0.40	-1.51 (1.04)	**56.45**
Language (verbal reasoning)					
Raw score (sum score)	28.9 (3.89)	25.3 (4.72)	**102.36**	20.0 (4.77)	**> 1000**
z-score	0.93 (0.64)	0.36 (0.84)	**41.68**	-0.37 (0.69)	**119.41**
Visuo-spatial function					
Raw score (sum score)	12.5 (2.16)	11.9 (2.47)	0.40	10.5 (2.73)	2.39
z-score	0.14 (0.92)	-0.09 (1.07)	0.37	-0.70 (1.19)	2.56
Impaired cognitive measures n(%)					
0	22 (56.4%)	11 (22.0%)		3 (11.5%)	
1	14 (35.9%)	23 (46.0%)		8 (30.8%)	
2	2 (5.1%)	10 (20.0%)		6 (23.1%)	
3	1 (2.6%)	5 (10.0%)		8 (30.8%)	
4	0	1 (2.0%)		0	
5	0	0		1 (3.8%)	

Notes. Two-sided independent samples Bayesian *t*-tests with $r = .707$ and JZS prior distribution for continuous variables. Impairment is defined as a *z*-score ≤ -1.5. Due to a lack of norm values for the verbal working memory task, this task is not considered in the number of impaired cognitive measures.

Table S4.

Results from sociodemographic and clinical questionnaires per group.

	HC	PD-NC	HC vs. PD-NC BF_{10}	PD-MCI	PD-NC vs. PD-MCI BF_{10}
Education and profession					
Professional training			**5.49**		1.69
None	1 (2.5%)	2 (4.0%)		4 (13.8%)	
Apprenticeship	7 (17.5%)	22 (44.0%)		18 (62.1%)	
Applied studies	8 (20.0%)	12 (24.0%)		5 (17.2%)	
University studies	20 (50.0%)	11 (22.0%)		1 (3.4%)	
Post-gradual qualifications	4 (10.0%)	3 (6.0%)		1 (3.4%)	
Job-related math experience			0.06		**9.10**
Not at all	4 (10.0%)	7 (14.0%)		3 (10.3%)	
Slightly	8 (20.0%)	10 (20.0%)		2 (6.9%)	
Average	15 (37.5%)	17 (34.0%)		5 (17.2%)	
Considerable	8 (20.0%)	11 (22.0%)		17 (58.6%)	
Extraordinary	5 (12.5%)	5 (10.0%)		2 (6.9%)	
Interest in mathematics			0.09		0.18
Not at all	5 (12.5%)	7 (14.0%)		2 (6.9%)	
Slightly	8 (20.0%)	9 (18.0%)		3 (10.3%)	
Average	13 (32.5%)	20 (40.0%)		13 (44.8%)	
Considerable	11 (27.5%)	9 (18.0%)		9 (31.0%)	
Extraordinary	3 (7.5%)	5 (10.0%)		2 (6.9%)	
Subjective usefulness of math			0.03		0.73
Not at all	1 (2.5%)	2 (4.0%)		1 (3.4%)	
Slightly	1 (2.5%)	3 (6.0%)		3 (10.3%)	
Average	15 (37.5%)	16 (32.0%)		6 (20.7%)	
Considerable	19 (47.5%)	24 (48.0%)		10 (34.5%)	
Extraordinary	4 (10.0%)	5 (10.0%)		9 (31.0%)	
Level of income			0.26		0.29
Below average	2 (5.0%)	6 (12.0%)		1 (3.4%)	
Average	17 (42.5%)	22 (44.0%)		15 (51.7%)	
Above average	21 (52.5%)	22 (44.0%)		13 (44.8%)	
Orientation and organisation					
Paying bills on time (memory)			0.09		0.20
Not at all	0	0		0	
Slightly	0	0		1 (3.4%)	
Average	1 (2.5%)	2 (4.0%)		3 (10.3%)	
Considerable	12 (30.0%)	14 (28.0%)		11 (37.9%)	
Extraordinary	27 (67.5%)	34 (68.0%)		14 (48.3%)	
Organised official documents			0.02		0.03
Not at all	0	1 (2.0%)		0	
Slightly	0	1 (2.0%)		0	
Average	4 (10.0%)	8 (16.0%)		6 (20.7%)	
Considerable	15 (37.5%)	15 (30.0%)		12 (41.4%)	
Extraordinary	21 (52.5%)	25 (50.0%)		11 (37.9%)	
Financial risk-taking			0.04		**3.19**
Not at all	16 (40.0%)	18 (36.0%)		8 (27.6%)	
Slightly	14 (35.0%)	20 (40.0%)		9 (31.0%)	
Average	8 (20.0%)	9 (18.0%)		4 (13.8%)	
Considerable	2 (5.0%)	1 (2.0%)		7 (24.1%)	
Extraordinary	0	2 (4.0%)		1 (3.4%)	
Clinical					

Non-motor symptoms		8.09 (4.60)		10.8 (5.64)	2.23
Health-related quality of life		24.6 (19.9)		38.8 (32.9)	2.27
ADL participant rating	0.0513 (0.223)	1.65 (2.95)	**27.39**	5.10 (6.03)	**27.70**
ADL caregiver rating	0.118 (0.686)	2.00 (3.33)	**18.31**	6.96 (7.26)	**71.45**

Notes. Two-sided independent samples Bayesian *t*-tests with $r = 0.707$ and JZS prior distribution for continuous variables, and Bayesian contingency tables with joint multinomial sampling for categorical variables.

Supplementary Material D: Complete ANCOVA tables for research question 1

Table S5

Bayesian ANCOVA on addition performance.

		r scale		
Accuracy		1	0.5	2
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} complexity	**125.91**	**174.16**	**71.61**
	BF_{10} group + complexity	**92.09**	**219.32**	**27.17**
	BF_{10} education + complexity	**85.61**	**33.53**	**13.60**
	BF_{10} depression + group + complexity	**48.32**	**96.85**	**14.26**
	BF_{10} motor symptoms + group + complexity	**44.88**	**93.69**	**14.64**
	BF_{10} education + group + complexity	**28.87**	**74.37**	**9.13**
	BF_{10} motor symptoms + complexity	**28.30**	**38.62**	**16.72**
	BF_{10} education + motor symptoms + group + complexity	**23.49**	**31.07**	**4.97**
	BF_{10} depression + complexity	**22.84**	**31.96**	**12.26**
Analysis of effects	BF_{incl} participant	*0.10*	*0.10*	*0.12*
	BF_{incl} education	0.45	0.36	*0.29*[#]
	BF_{incl} motor symptoms	0.36	0.38	0.39
	BF_{incl} depression	0.34	0.45	0.34
	BF_{incl} group	0.96	1.70	0.63
	BF_{incl} complexity	**172.69**	**222.27**	**85.27**
	BF_{incl} group × complexity	*0.17*	0.34[#]	*0.10*
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + LEDD + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + LEDD + disease stage + group × complexity	**> 1000**	**> 1000**	**706.46**
Analysis of effects	BF_{incl} participant	*0.04*	*0.03*	*0.04*
	BF_{incl} LEDD	*0.20*	*0.19*	*0.20*
	BF_{incl} Hoehn & Yahr disease stage	*0.26*	*0.24*	*0.24*
	BF_{incl} group	**66.69**	**94.18**	**38.38**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.25*	0.41[#]	*0.13*
Reaction Time				
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + education + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms + depression + group × complexity	**> 1000**	**> 1000**	**> 1000**

	BF_{10} complexity + participant + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + depression + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + group × complexity	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} education	**3.38**	1.87[#]	**3.32**
	BF_{incl} motor symptoms	**5.93**	**3.50**	**3.77**
	BF_{incl} depression	*0.16*	*0.26*	0.65[#]
	BF_{incl} group	2.25	2.71	**4.08**[#]
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	0.94	1.63	0.52
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + group × complexity	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} LEDD	0.33	0.27	0.28
	BF_{incl} Hoehn & Yahr disease stage	**45.50**	**57.64**	**28.45**
	BF_{incl} group	1.60	0.77	0.95
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	0.78	1.86	*0.26*[#]

Notes. Models were compared to the null model and effects were estimated across matched models. For model comparisons, the best 10 models are shown and the order stems from the size of BF_{10} for scale = 1. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were arcine-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. BFs > 3 were marked in bold and provide evidence in favour of including the respective variable. BFs < 1/3 were marked in italics and provide evidence against including the respective variable. BFs that were not marked need to be interpreted as inconclusive.

Table S6

Bayesian ANCOVA on subtraction performance

		r scale		
Accuracy		1	0.5	2
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant +education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} education	0.87	0.91	0.92
	BF_{incl} motor symptoms	0.53	0.50	0.54
	BF_{incl} depression	0.38	0.35	0.41
	BF_{incl} group	1.54	2.02	0.81
	BF_{incl} complexity	**34.91**	**50.33**	**20.49**
	BF_{incl} group × complexity	0.39	0.60	*0.19*[#]
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} LEDD	0.53	0.54	0.53
	BF_{incl} Hoehn & Yahr disease stage	0.37	0.36	0.38
	BF_{incl} group	**15.14**	**16.59**	**9.98**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.31*	0.43[#]	*0.14*
Reaction Time				
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**

	BF_{10} group + complexity + participant + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms + depression + group × complexity	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} education	**21.48**	**23.44**	**20.02**
	BF_{incl} motor symptoms	**3.04**	**3.63**	1.91[#]
	BF_{incl} depression	0.33	*0.32*[#]	0.33
	BF_{incl} group	**61.15**	**37.74**	**74.24**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.03*	*0.09*	*0.01*
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + group × complexity	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} LEDD	0.34	0.35	0.38
	BF_{incl} Hoehn & Yahr disease stage	**30.88**	**38.02**	**18.68**
	BF_{incl} group	**38.94**	**23.24**	**34.41**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.03*	*0.11*	*0.01*

Notes. Models were compared to the null model and effects were estimated across matched models. For model comparisons, the best 10 models are shown and the order stems from the size of BF_{10} for scale = 1. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were arcine-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. BFs > 3 were marked in bold and provide evidence in favour of including the respective variable. BFs < 1/3 were marked in italics and provide evidence against including the respective variable. BFs that were not marked need to be interpreted as inconclusive.

Table S7

Bayesian ANCOVA on multiplication performance

		r scale		
Accuracy		1	0.5	2
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} complexity + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + education + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + education + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	*0.06*	*0.06*	*0.06*
	BF_{incl} education	0.69	0.65	0.70
	BF_{incl} motor symptoms	0.91	0.85	0.96
	BF_{incl} depression	0.40	0.39	0.43
	BF_{incl} group	*0.18*	0.35[#]	*0.09*
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.14*	*0.28*	*0.08*
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + disease stage + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	*0.26*	*0.23*	*0.27*
	BF_{incl} LEDD	*0.22*	*0.22*	*0.22*
	BF_{incl} Hoehn & Yahr disease stage	0.81	0.77	0.79
	BF_{incl} group	*0.12*	*0.24*	*0.07*
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.13*	*0.23*	*0.07*
Reaction Time				
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms +	**> 1000**	**> 1000**	**> 1000**

	group × complexity			
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} education	**3.12**	**3.65**	**3.68**
	BF_{incl} motor symptoms	**10.92**	**13.07**	**7.46**
	BF_{incl} depression	0.31	0.29	0.33
	BF_{incl} group	**5.74**	**5.93**	**3.28**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.08*	*0.25*	*0.02*
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + group × complexity	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} LEDD	*0.22*	*0.12*	*0.24*
	BF_{incl} Hoehn & Yahr disease stage	**147.90**	**345.14**	**131.80**
	BF_{incl} group	**12.85**	**27.75**	**9.05**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.05*	*0.10*	*0.02*

Notes. Models were compared to the null model and effects were estimated across matched models. For model comparisons, the best 10 models are shown and the order stems from the size of BF_{10} for scale = 1. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were logit-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. BFs > 3 were marked in bold and provide evidence in favour of including the respective variable. BFs < 1/3 were marked in italics and provide evidence against including the respective variable. BFs that were not marked need to be interpreted as inconclusive.

Table S8.

Bayesian ANCOVA on division performance

		r scale		
Accuracy		1	0.5	2
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} complexity + participant + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education + motor symptoms + depression	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	1.96	1.80	2.04
	BF_{incl} education	1.05	1.03	1.02
	BF_{incl} motor symptoms	**3.37**	**3.29**	**3.59**
	BF_{incl} depression	*0.30*	*0.30*	*0.31*
	BF_{incl} group	*0.20*	0.35[#]	*0.11*
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.12*	*0.23*	*0.06*
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} complexity + disease stage	**357.44**	**516.96**	**200.97**
	BF_{10} complexity + participant + disease stage	**328.78**	**312.02**	**132.92**
	BF_{10} complexity + LEDD + disease stage	**131.47**	**183.73**	**73.41**
	BF_{10} complexity + participant + LEDD + disease stage	**92.62**	**125.35**	**52.52**
	BF_{10} group + complexity + disease stage	**54.49**	**142.31**	**15.75**
	BF_{10} group + complexity + participant + disease stage	**44.30**	**99.65**	**12.57**
	BF_{10} complexity + participant	**37.12**	**52.51**	**21.11**
	BF_{10} complexity	**23.78**	**34.63**	**13.24**
	BF_{10} complexity + participant + LEDD	**22.05**	**30.12**	**13.00**
Analysis of effects	BF_{incl} participant	0.91	0.73	0.72
	BF_{incl} LEDD	0.35	0.39	0.40
	BF_{incl} Hoehn & Yahr disease stage	**7.79**	**6.44**	**7.31**
	BF_{incl} group	*0.16*	*0.32*	*0.09*
	BF_{incl} complexity	**53.39**	**69.00**	**27.76**
	BF_{incl} group × complexity	*0.28*	0.45[#]	*0.13*
Reaction Time				
HC vs. PD-NC				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + depression + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + education + depression + motor symptoms	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + education + motor symptoms +	**> 1000**	**> 1000**	**> 1000**

	group × complexity			
	BF_{10} group + complexity + participant + depression + motor symptoms	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} education	**43.65**	**54.79**	**38.85**
	BF_{incl} motor symptoms	**113.61**	**195.63**	**74.10**
	BF_{incl} depression	0.24	0.20	0.22
	BF_{incl} group	**11.95**	**13.80**	**6.47**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.03*	*0.10*	*0.01*
PD-NC vs. PD-MCI				
Model comparison	BF_{10} Null model	1.00	1.00	1.00
	BF_{10} group + complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD + disease stage + group × complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{10} complexity + participant + LEDD + disease stage	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + LEDD	**> 1000**	**> 1000**	**> 1000**
	BF_{10} group + complexity + participant + group × complexity	**> 1000**	**> 1000**	**> 1000**
Analysis of effects	BF_{incl} participant	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} LEDD	*0.19*	*0.20*	*0.17*
	BF_{incl} Hoehn & Yahr disease stage	**> 1000**	**> 1000**	**838.58**
	BF_{incl} group	**56.93**	**47.15**	**38.73**
	BF_{incl} complexity	**> 1000**	**> 1000**	**> 1000**
	BF_{incl} group × complexity	*0.03*	*0.11*	*0.01*

Notes. Models were compared to the null model and effects were estimated across matched models. For model comparisons, the best 10 models are shown and the order stems from the size of BF_{10} for scale = 1. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were logit-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. BFs > 3 were marked in bold and provide evidence in favour of including the respective variable. BFs < 1/3 were marked in italics and provide evidence against including the respective variable. BFs that were not marked need to be interpreted as inconclusive.

Supplementary Material E: ANCOVA tables for research question 2

Table S9.

Bayesian ANCOVA on addition performance with domain-general variables

		BF_{incl} for r scale		
Accuracy		1	0.5	2
HC vs. PD-NC	participant	*0.10*	*0.09*	*0.11*
	education	0.44	0.45	0.41
	motor symptoms	0.52	0.50	0.49
	depression	0.38	0.38	0.35
	group	1.41	2.19	0.78
	complexity	**165.85**	**228.77**	**91.30**
	group × complexity	*0.18*	0.34[#]	*0.09*
	verbal working memory	0.99	0.96	0.94
PD-NC vs. PD-MCI	participant	*0.04*	*0.04*	*0.04*
	LEDD	*0.28*	*0.28*	*0.30*
	Hoehn & Yahr disease stage	*0.29*	*0.31*	*0.31*
	group	**4.90**	**7.04**	**2.89[#]**
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	0.36	0.54	*0.19[#]*
	verbal working memory	1.42	1.45	1.56
Reaction Time				
HC vs. PD-NC	participant	**> 1000**	**> 1000**	**> 1000**
	education	**4.09**	2.52[#]	**5.67**
	motor symptoms	2.60	2.39	**4.12[#]**
	depression	*0.24*	*0.31*	0.23
	group	2.84	2.65	**4.08[#]**
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	0.94	1.08	*0.11[#]*
	visuo-spatial working memory	**42.89**	**57.27**	**81.68**
PD-NC vs. PD-MCI	participant	**> 1000**	**> 1000**	**> 1000**
	LEDD	0.27	0.25	0.31
	Hoehn & Yahr disease stage	**11.54**	**14.93**	**6.67**
	group	1.36	1.43	1.13
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	0.84	1.28	*0.21[#]*
	visuo-spatial working memory	**50.95**	**65.90**	**57.51**

Notes. Models were compared to the null model and effects were estimated across matched models. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were arcine-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. Values marked in bold are larger than BF = 3 and provide evidence in favour of including the respective variable. Values marked in italics are smaller than BF = 1/3 and provide evidence against including the respective variable. Those values not marked need to be interpreted as inconclusive.

Table S10.

Bayesian ANCOVA on subtraction performance with domain-general variables

		BF_{incl} for r scale		
Accuracy		1	0.5	2
HC vs. PD-NC	participant	**> 1000**	**> 1000**	**> 1000**
	education	0.49	0.53	0.50
	motor symptoms	0.57	0.59	0.50
	depression	0.36	0.38	0.36
	group	0.77	1.10	0.45
	complexity	**33.75**	**49.69**	**19.38**
	group × complexity	0.38	0.69	0.19
	verbal reasoning/ language	**4.61**	**4.40**	**5.00**
PD-NC vs. PD-MCI	participant	**> 1000**	**> 1000**	**> 1000**
	LEDD	0.50	0.51	0.53
	Hoehn & Yahr disease stage	0.48	0.49	0.46
	group	**11.99**	**12.46**	**8.29**
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	0.40	0.63	*0.22*[#]
	verbal reasoning/ language	0.96	1.08	1.00
Reaction Time				
HC vs. PD-NC	participant	**> 1000**	**> 1000**	**> 1000**
	education	**5.92**	**5.82**	**5.60**
	motor symptoms	1.77	2.06	1.49
	depression	0.39	0.37	0.36
	group	**19.90**	**13.64**	**21.04**
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.03*	*0.09*	*0.01*
	verbal reasoning/ language	1.57	1.89	1.59
PD-NC vs. PD-MCI	participant	**> 1000**	**> 1000**	**> 1000**
	LEDD	0.34	0.33	0.79
	Hoehn & Yahr disease stage	**5.20**	**6.88**	**5.48**
	group	**7.15**	**5.47**	**9.20**
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.03*	0.10	*0.01*
	verbal reasoning/ language	2.27	2.40	2.50

Notes. Models were compared to the null model and effects were estimated across matched models. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were arcine-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. Values marked in bold are larger than BF = 3 and provide evidence in favour of including the respective variable. Values marked in italics are smaller than BF = 1/3 and provide evidence against including the respective variable. Those values not marked need to be interpreted as inconclusive.

Table S11.

Bayesian ANCOVA on multiplication performance with domain-general variables

		BF_{incl} for r scale		
Accuracy		1	0.5	2
HC vs. PD-NC	participant	*0.06*	*0.05*	*0.06*
	education	0.56	0.56	0.58
	motor symptoms	0.74	0.71	0.77
	depression	0.40	0.39	0.41
	group	*0.17*	*0.30*	*0.09*
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.15*	*0.29*	*0.08*
	verbal reasoning/ language	0.68	0.64	0.68
PD-NC vs. PD-MCI	participant	*0.15*	*0.15*	*0.16*
	LEDD	*0.28*	*0.29*	*0.27*
	Hoehn & Yahr disease stage	1.31	1.32	1.37
	group	*0.12*	*0.22*	*0.06*
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.13*	*0.24*	*0.07*
	verbal reasoning/ language	1.83	1.76	1.85
Reaction Time				
HC vs. PD-NC	participant	**> 1000**	**> 1000**	**> 1000**
	education	1.19	1.38	0.92
	motor symptoms	2.24	2.11	1.97
	depression	0.42	0.50	0.55
	group	1.57	2.57	0.98
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.09*	*0.23*	*0.03*
	verbal working memory	2.93	**3.06**[#]	2.63
PD-NC vs. PD-MCI	participant	**> 1000**	**> 1000**	**> 1000**
	LEDD	0.25	0.29	0.28
	Hoehn & Yahr disease stage	**13.20**	**13.71**	**13.00**
	group	0.95	1.18	*0.27*[#]
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.05*	*0.22*	*0.02*
	verbal working memory	**24.37**	**25.13**	**27.06**

Notes. Models were compared to the null model and effects were estimated across matched models. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were logit-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. Values marked in bold are larger than BF = 3 and provide evidence in favour of including the respective variable. Values marked in italics are smaller than BF = 1/3 and provide evidence against including the respective variable. Those values not marked need to be interpreted as inconclusive.

Table S12.

Bayesian ANCOVA on division performance with domain-general variables

		BF_{incl} for r scale		
Accuracy		1	0.5	2
HC vs. PD-NC	participant	1.59	1.41	1.57
	education	0.49	0.56	0.57
	motor symptoms	1.36	1.37	1.35
	depression	*0.27*	*0.30*	*0.31*
	group	*0.15*	*0.32*	*0.09*
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.13*	*0.23*	*0.06*
	verbal reasoning/ language	3.00	2.61	2.67
PD-NC vs. PD-MCI	participant	0.34	0.34	0.34
	LEDD	*0.33*	*0.33*	*0.33*
	Hoehn & Yahr disease stage	0.77	0.74	0.72
	group	*0.12*	*0.22*	*0.06*
	complexity	**69.74**	**98.13**	**39.93**
	group × complexity	*0.23*	0.41[#]	*0.12*
	verbal reasoning/ language	**18.50**	**19.03**	**20.35**
Reaction Time				
HC vs. PD-NC	participant	**> 1000**	**> 1000**	**> 1000**
	education	**5.75**	**5.79**	**5.42**
	motor symptoms	**19.53**	**30.05**	**14.83**
	depression	0.27	0.25	0.29
	group	**3.50**	**3.88**	2.05[#]
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.03*	*0.10*	*0.01*
	verbal reasoning/ language	1.56	1.78	1.69
PD-NC vs. PD-MCI	participant	**> 1000**	**> 1000**	**> 1000**
	LEDD	0.22	0.40	0.19
	Hoehn & Yahr disease stage	**39.56**	**62.87**	**33.02**
	group	**3.98**	**6.27**	**3.19**
	complexity	**> 1000**	**> 1000**	**> 1000**
	group × complexity	*0.05*	*0.08*	*0.01*
	verbal reasoning/ language	1.93	**3.43**	1.97

Notes. Models were compared to the null model and effects were estimated across matched models. Group and complexity were treated as fixed factors, while subjects were treated as random factors. The dependent variables were logit-transformed accuracy data and log-transformed reaction time data. [#] The robustness check yields a deviating result. Values marked in bold are larger than BF = 3 and provide evidence in favour of including the respective variable. Values marked in italics are smaller than BF = 1/3 and provide evidence against including the respective variable. Those values not marked need to be interpreted as inconclusive.

Supplementary Material F: Exploration of correlations per group

Table S13.

Pearson correlations between arithmetic performance and cognitive domains in the overall sample

	Addition *N* = 115		Subtraction *N* = 114		Multiplication *N* = 116		Division *N* = 116	
	ACC	RT	ACC	RT	ACC	RT	ACC	RT
Executive functions	*r* = -.08 [-.26; .10] *n* = 113	*r* = .14 [-.04; .31] *n* = 113	*r* = -.07 [-.25; .12] *n* = 112	*r* = .10 [-.09; .28] *n* = 112	*r* = -.28* [-.43; -.10] *n* = 115	*r* =-.10 [-.09; .27] *n* = 115	*r* = -.12 [-.30; .06] *n* = 115	*r* = .12 [-.07; .29] *n* = 115
Verbal working memory	***r* = .38*** ***[.21; .52]*** ***n* = 114**	*r* = -.39* [-.53; -.22] *n* = 114	*r* = .24* [.06; .40] *n* = 113	*r* = -.37* [-.51; -.19] *n* = 113	*r* = .17 [-.02; .34] *n* = 115	***r* = -.33*** ***[-.47; -.15]*** ***n* = 115**	*r* = .27* [.09; .42] *n* = 115	*r* = -.29* [-.44; -.11] *n* = 115
Visuo-spatial working memory	*r* = .26* [.08; .42] *n* = 115	***r* = -.39*** ***[-.53; -.22]*** ***n* = 115**	*r* = .35* [.17; .49] *n* = 114	*r* = -.36* [-.51; -.19] *n* = 114	*r* = .18* [.00; .35] *n* = 116	*r* = -.14 [-.31; .05] *n* = 116	*r* = .19* [.00; .35] *n* = 116	*r* = -.24* [-.40; -.06] *n* = 116
Attention	*r* = -.12 [-.30; .06] *n* = 113	*r* = .33* [.15; .48] *n* = 113	*r* = -.24* [-.40; -.05] *n* = 112	*r* = .31* [.13; .47] *n* = 112	*r* = -.11 [-.29; .07] *n* = 115	*r* = .20* [.01; .36] *n* = 115	*r* = -.06 [-.24; .12] *n* = 115	*r* = .24* [.05; .40] *n* = 115
Verbal memory	*r* = .24* [.05; .40] *n* = 112	*r* = -.18 [-.35; .01] *n* = 112	*r* = .33* [.15; .48] *n* = 111	*r* = -.17 [-.35; .01] *n* = 111	*r* = .15 [-.03; .32] *n* = 113	*r* = -.09 [-.26; .10] *n* = 113	*r* = .32* [.14; .47] *n* = 113	*r* = -.24* [-.40; -.05] *n* = 113
Language/ verbal reasoning	*r* = .25* [.06; .41] *n* = 114	*r* = -.34* [-.49; -.17] *n* = 114	***r* = .44*** ***[.28; .58]*** ***n* = 113**	***r* = -.44*** ***[-.57; -.27]*** ***n* = 113**	***r* = .31*** ***[.13; .46]*** ***n* = 115**	*r* = -.25* [-.41; -.07] *n* = 115	***r* = .54*** ***[.39; .65]*** ***n* = 115**	***r* = -.47*** ***[-.60; -.31]*** ***n* = 115**
Visuo-spatial function	*r* = .22* [.03; .38] *n* = 115	*r* = -.28* [-.43; -.10] *n* = 115	*r* = .19* [.01; .36] *n* = 114	*r* = -.33* [-.48; -.16] *n* = 114	*r* = .16 [-.03; .32] *n* = 116	*r* = -.32* [-.47; -.15] *n* = 116	*r* = .35* [.18; .50] *n* = 116	*r* = -.33* [-.48; -.15] *n* = 116

Note. N = overall sample size indicating the number of participants in all three groups solving the task with sufficient accuracy to ensure task understanding. Values marked as bold indicate the correlation of highest magnitude for the respective arithmetic ask. * $p < .05$.

Table S14.

Pearson correlations between arithmetic performance and cognitive domains in healthy controls

	Addition		Subtraction		Multiplication		Division	
	ACC	RT	ACC	RT	ACC	RT	ACC	RT
Executive functions	*r* = -.12 *n* = 38	*r* = .09 *n* = 38	***r* = -.30** ***n* = 39**	*r* = .27 *n* = 39	*r* = -.07 *n* = 39	*r* = -.21 *n* = 39	*r* = -.07 *n* = 39	*r* = .28 *n* = 39
Verbal working memory	*r* = .19 *n* = 39	***r* = -.36*** ***n* = 39**	*r* = .21 *n* = 40	***r* = -.34*** ***n* = 40**	*r* = .01 *n* = 39	*r* = -.11 *n* = 39	***r* = .24** ***n* = 39**	*r* = -.07 *n* = 39
Visuo-spatial working memory	*r* = .06 *n* = 39	*r* = -.20 *n* = 39	*r* = .07 *n* = 40	*r* = -.15 *n* = 40	*r* = -.05 *n* = 39	*r* = -.29 *n* = 39	*r* = .24 *n* = 39	*r* = -.16 *n* = 39
Attention	***r* = -.22** ***n* = 38**	*r* = .27* *n* = 38	*r* = .07 *n* = 39	*r* = .30 *n* = 39	*r* = .04 *n* = 39	***r* = .26** ***n* = 39**	*r* = .03 *n* = 39	*r* = .04 *n* = 39
Verbal memory	*r* = -.01 *n* = 39	*r* = -.10 *n* = 39	*r* = -.05 *n* = 40	*r* = -.15 *n* = 40	***r* = -.33** ***n* = 39**	*r* = .14 *n* = 39	*r* = .10 *n* = 39	*r* = .19 *n* = 39

Language/ verbal reasoning	r = -.06 n = 39	r = -.08 n = 39	r = .29 n = 40	r = -.18 n = 40	r = -.07 n = 39	r = .02 n = 39	r = .20 n = 39	**r = -.15** **n = 39**
Visuo-spatial function	r = .10 n = 39	r = -.08 n = 39	r = .03 n = 40	r = -.16 n = 40	r = -.01 n = 39	r = -.21 n = 39	r = .23 n = 39	r = -.17 n = 39

Note. Values marked as bold indicate the correlation of highest magnitude for the respective arithmetic task. * $p < .05$. Values marked in yellow indicate different directions of correlations between groups.

Table S15.

Pearson correlations between arithmetic performance and cognitive domains in PD patients with normal cognition (PD-NC)

	Addition		Subtraction		Multiplication		Division	
	ACC	RT	ACC	RT	ACC	RT	ACC	RT
Executive functions	r = .02 n = 48	r = .04 n = 48	r = -.11 n = 47	r = -.02 n = 47	r = -.29* n = 49	r = -.02 n = 49	r = .02 n = 48	r = -.01 n = 48
Verbal working memory	**r = .27** **n = 48**	r = -.39* n = 48	r = .20 n = 47	r = -.38* n = 47	r = .23 n = 49	**r = -.32*** **n = 49**	r = .26* n = 48	r = -.32* n = 48
Visuo-spatial working memory	r = .21 n = 48	**r = -.44*** **n = 48**	**r = .34*** **n = 47**	**r = -.38*** **n = 47**	r = .29* n = 49	r = -.17 n = 49	r = .08 n = 48	r = -.24 n = 48
Attention	r = .18 n = 48	r = .17 n = 48	r = -.10 n = 47	r = .08 n = 47	r = -.10 n = 49	r = .05 n = 49	r = .07 n = 48	r = .12 n = 48
Verbal memory	r = .07 n = 48	r = -.31* n = 48	r = .26 n = 47	r = -.31* n = 47	**r = .43*** **n = 49**	r = -.26 n = 49	r = .36* n = 48	r = -.42* n = 48
Language/ verbal reasoning	r = -.06 n = 48	r = -.22 n = 48	r = .26 n = 47	r = -.32 n = 47	r = .41* n = 49	r = -.08 n = 49	**r = .47*** **n = 48**	**r = -.46*** **n = 48**
Visuo-spatial function	r = .17 n = 48	r = -.16 n = 48	r = .12 n = 47	r = -.13 n = 47	r = .16 n = 49	r = -.30* n = 49	r = .25* n = 48	r = -.28* n = 48

Note. Values marked as bold indicate the correlation of highest magnitude for the respective arithmetic task. * $p < .05$. Values marked in yellow indicate different directions of correlations between groups.

Table S16.

Pearson correlations between arithmetic performance and cognitive domains in PD patients with Mild Cognitive impairment (PD-MCI)

	Addition		Subtraction		Multiplication		Division	
	ACC	RT	ACC	RT	ACC	RT	ACC	RT
Executive functions	r = .08 n = 27	r = .07 n = 27	r = .29 n = 26	r = -.01 n = 26	r = -.21 n = 27	r = .23 n = 27	r = -.18 n = 28	r = .03 n = 28
Verbal working memory	r = .19 n = 27	r = -.12* n = 27	r = -.20 n = 26	r = .01 n = 26	r = -.06 n = 27	r = -.20 n = 27	r = .07 n = 28	r = .07 n = 28
Visuo-spatial working	r = .17 n = 28	r = -.26 n = 28	**r = .39*** **n = 27**	r = -.27 n = 27	r = -.03 n = 28	r = .22 n = 28	r = .13 n = 29	r = .03 n = 29

memory								
Attention	***r* = -.07** ***n* = 27**	*r* = .36 *n* = 27	*r* = -.30 *n* = 26	*r* = .36 *n* = 26	*r* = -.05 *n* = 27	*r* = .06 *n* = 27	*r* = -.02 *n* = 28	***r* = .28** ***n* = 28**
Verbal memory	*r* = -.01 *n* = 25	***r* = .53*** ***n* = 25**	*r* = .17 *n* = 24	***r* = .62*** ***n* = 24**	***r* = -.18** ***n* = 25**	***r* = .43*** ***n* = 25**	*r* = .19 *n* = 26	*r* = .27 *n* = 26
Language/ verbal reasoning	*r* = -.17 *n* = 27	*r* = -.04 *n* = 27	*r* = .12 *n* = 26	*r* = -.09 *n* = 26	*r* = .02 *n* = 27	*r* = -.18 *n* = 27	***r* = .57*** ***n* = 28**	*r* = -.08 *n* = 28
Visuo-spatial function	*r* = -.03 *n* = 28	*r* = -.31 *n* = 28	*r* = -.02 *n* = 27	*r* = -.32 *n* = 27	*r* = .06 *n* = 28	*r* = -.26 *n* = 28	*r* = .36 *n* = 29	*r* = -.18 *n* = 29

Note. Values marked as bold indicate the correlation of highest magnitude for the respective arithmetic task. * $p < .05$. Values marked in yellow indicate different directions of correlations between groups.

Table S17.

Pearson correlations between accuracy and reaction time of each arithmetic operation per group

		RT Addition	RT Subtraction	RT Multiplication	RT Division
ACC Addition	HC	-.20 [-.48; .13]			
	PD-NC	-.24 [-.49; .05]			
	PD-MCI	-.03 [-.40; .34]			
ACC Subtraction	HC		-.32* [-.57; -.01]		
	PD-NC		-.35* [-.58; -.07]		
	PD-MCI		-.14 [-.49; .25]		
ACC Multiplication	HC			-.49* [-.70; -.20]	
	PD-NC			-.42* [-.62; -.15]	
	PD-MCI			-.19 [-.53; .19]	
ACC Division	HC				-.24 [-.52; .08]
	PD-NC				-.46* [-.66; -.21]
	PD-MCI				-.39* [-.66; -.03]

Note. * $p < .05$.

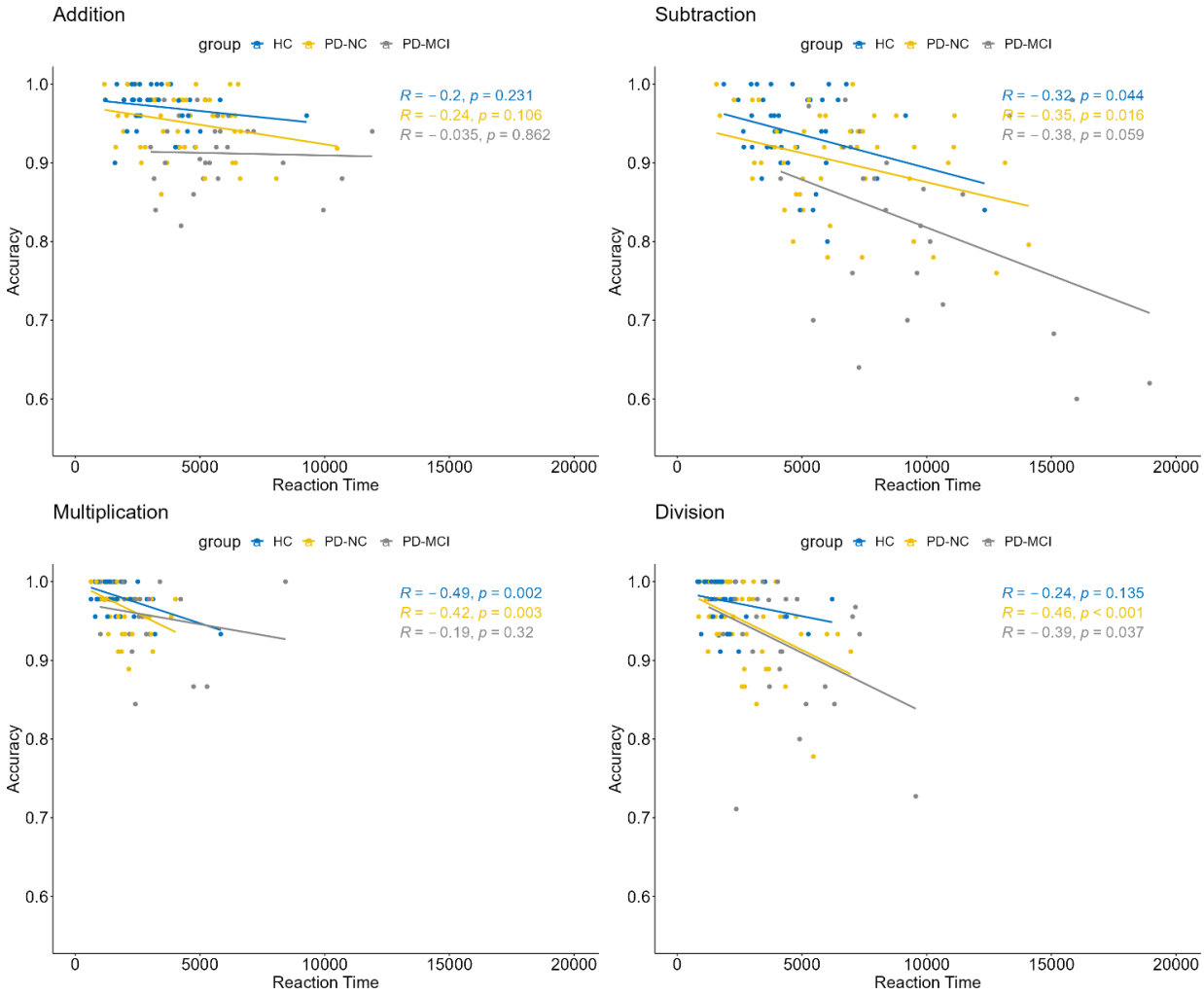

Figure S1. Association between accuracy and reaction time per arithmetic task and group.

Supplementary Material G: Evidence accumulation plots for research question 1

Comparing the evidence accumulation for addition reaction time between HC and PD-NC to the main ANCOVA table indicates that evidence is instable. All other ANCOVAs yield stable evidence.

Addition

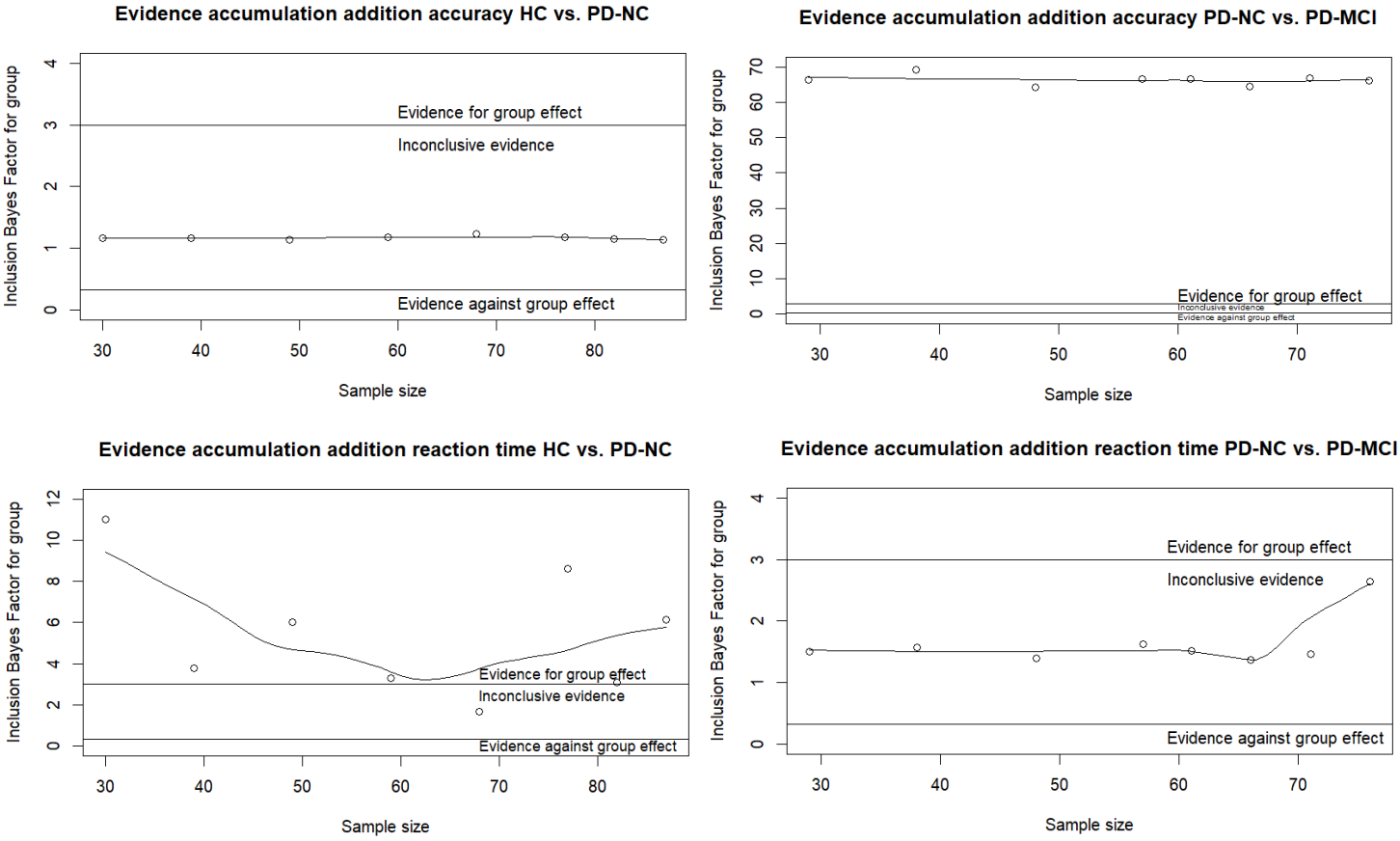

Figure S2. *Plotting evidence accumulation for the factor group in ANCOVAs for addition. The overall sample was tested in steps of n = 45, 60, 75, 89, 99, 109, 114, 119.*

Subtraction

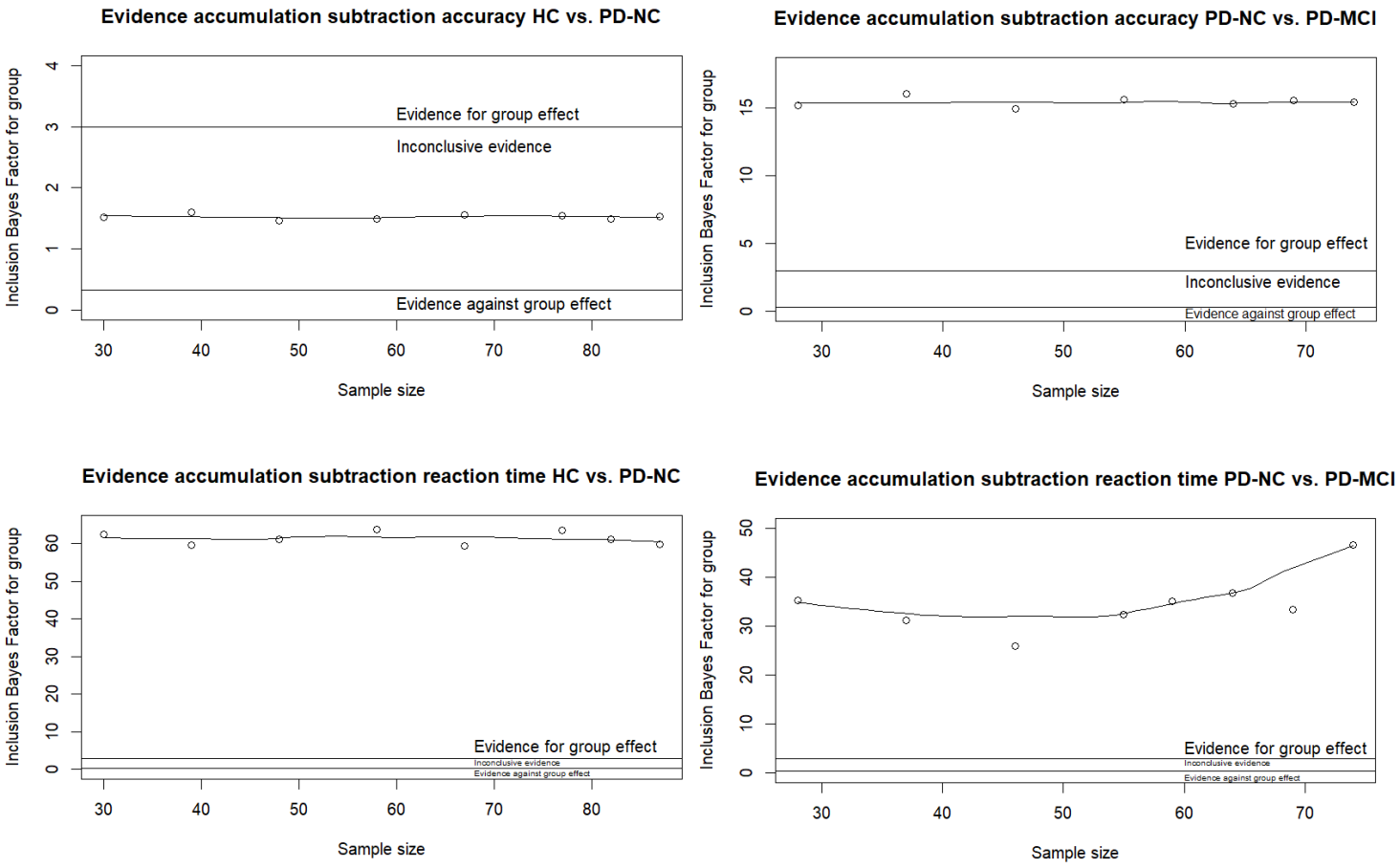

Figure S3. *Plotting evidence accumulation for the factor group in ANCOVAs for subtraction. The overall sample was tested in steps of n = 45, 60, 75, 89, 99, 109, 114, 119.*

Multiplication

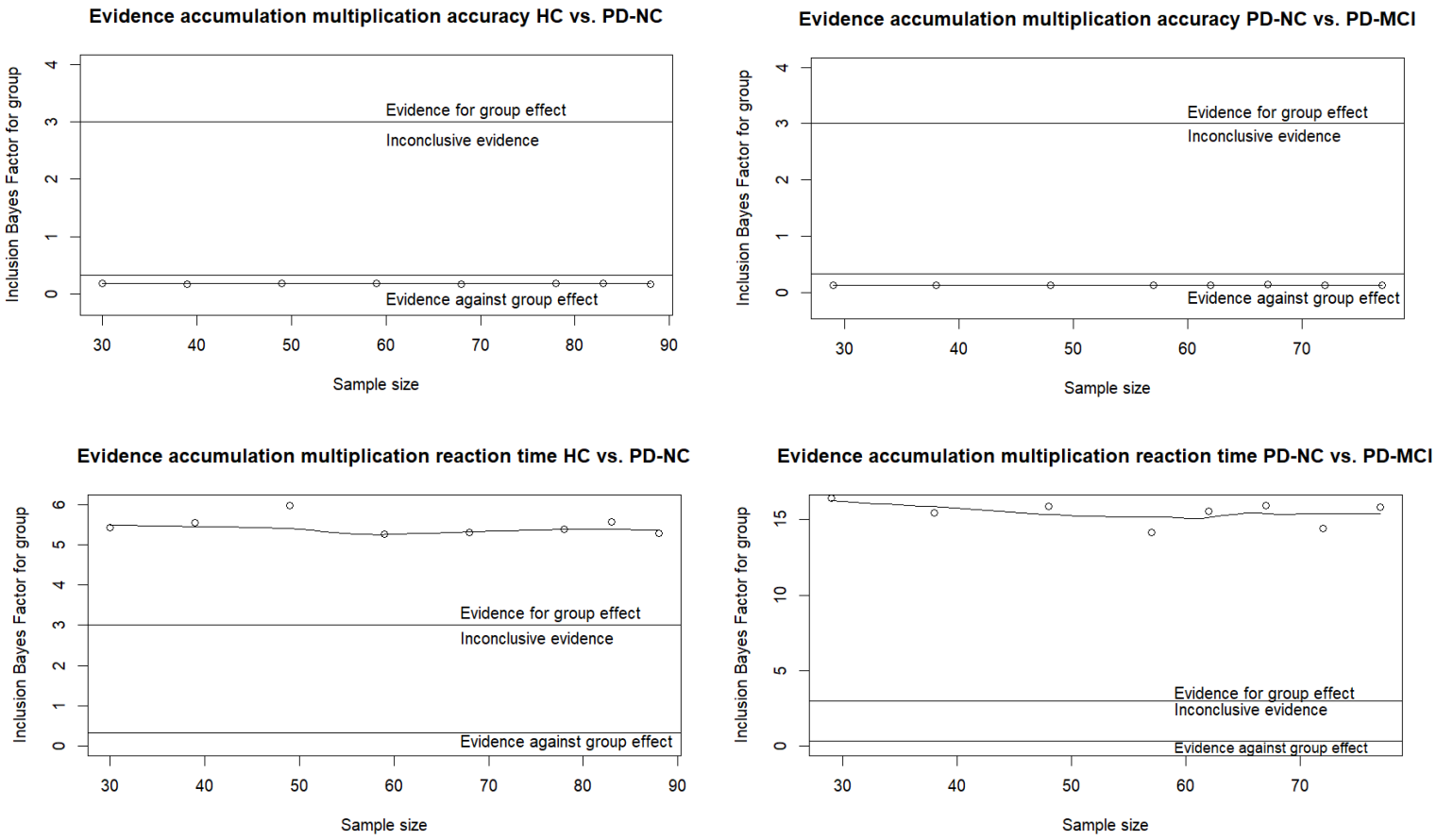

Figure S4. *Plotting evidence accumulation for the factor group in ANCOVAs for multiplication. The overall sample was tested in steps of n = 45, 60, 75, 89, 99, 109, 114, 119.*

Division

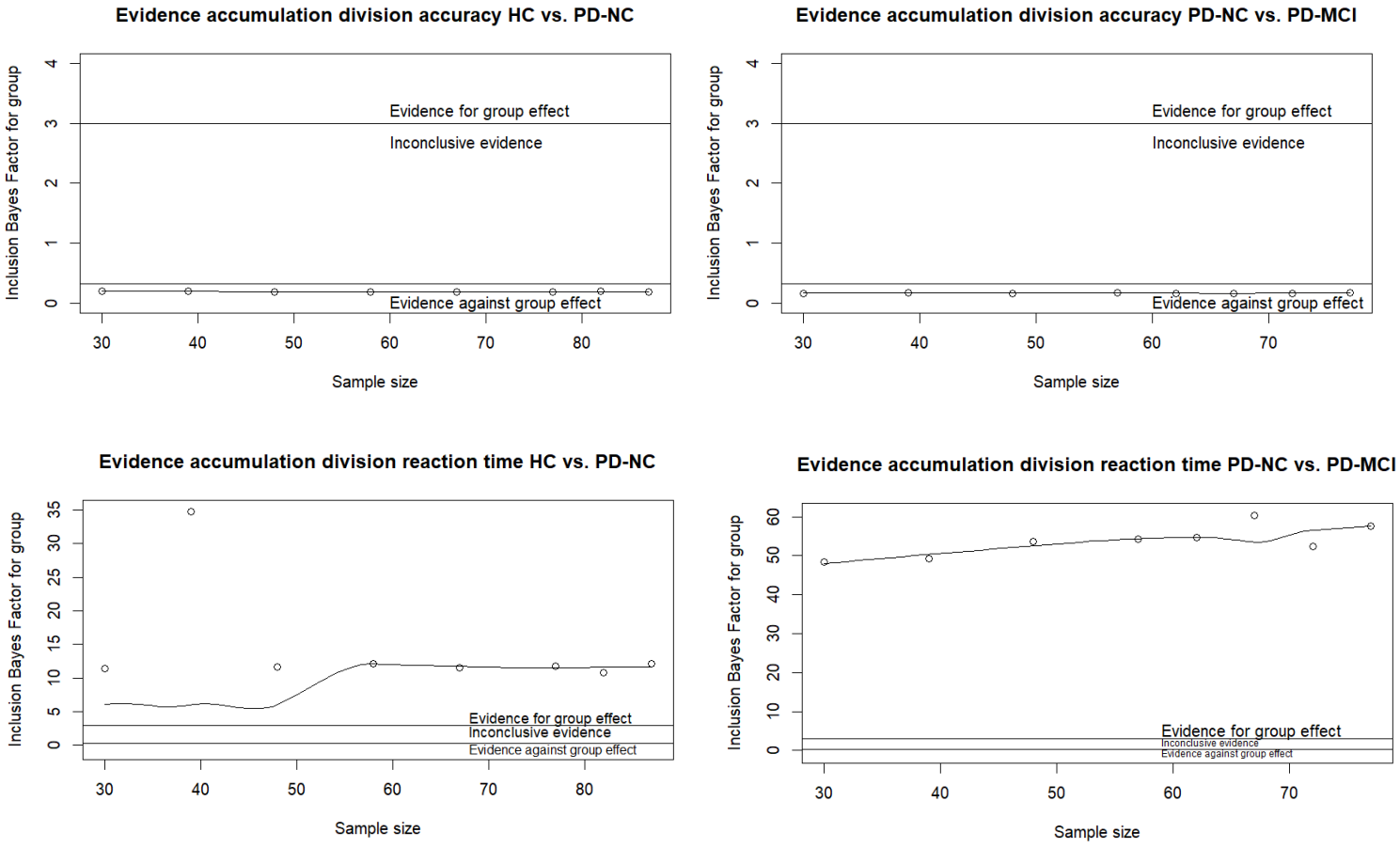

Figure S5. *Plotting evidence accumulation for the factor group in ANCOVAs for division. The overall sample was tested in steps of n = 45, 60, 75, 89, 99, 109, 114, 119.*