Received: 9 September 2020  |  Revised: 2 March 2021  |  Accepted: 12 March 2021
DOI: 10.1002/jnr.24907

ORIGINAL ARTICLE

Deficits in or preservation of basic number processing in Parkinson’s disease? A registered report

Hannah D. Loenneker1  |  Christina Artemenko1,2  |  Klaus Willmes3  |  Inga Liepelt-Scarfone4,5,6  |  Hans-Christoph Nuerk1

1Department of Psychology, University of Tuebingen, Tuebingen, Germany
2LEAD Graduate School & Research Network, University of Tuebingen, Tuebingen, Germany
3Department of Neurology, RWTH Aachen University, University Hospital, Aachen, Germany
4Department of Clinical Neurodegeneration, Hertie Institute for Clinical Brain Research, Tuebingen, Germany
5German Centre for Neurodegenerative Diseases, Tuebingen, Germany
6IB-Hochschule für Gesundheit und Soziales, Stuttgart, Germany

Correspondence
Hannah D. Loenneker, Department of Psychology, University of Tuebingen, Schleichstrasse 4, 72076 Tuebingen, Germany.
Email: hannah-dorothea.loenneker@uni-tuebingen.de

Funding information
Scholarship from Landesgraduiertenfoerderung Baden-Wuerttemberg and Friedrich-Ebert-Stiftung to HL. Grant over 3853.50€ for participants reimbursement and testing material from Wilhelm Schuler Foundation to HL.

Abstract
Neurodegenerative diseases such as Parkinson's disease (PD) have a huge impact on patients, caregivers, and the health-care system. To date, the diagnosis of mild cognitive impairments in PD has been established based on domain-general functions such as executive functions, attention, or working memory. However, specific numerical deficits observed in clinical practice have not yet been systematically investigated. PD-immanent deterioration of domain-general functions and domain-specific numerical areas suggests the mechanisms of both primary and secondary dyscalculia. The current study will systematically investigate basic number processing performance in PD patients for the first time, targeting domain-specific cognitive representations of numerosity and the influence of domain-general factors. The overall sample consists of patients with a diagnosis of PD, according to consensus guidelines, and healthy controls. PD patients will be stratified into patients with normal cognition or mild cognitive impairment (level I-PD-MCI based on cognitive screening). Basic number processing will be assessed using transcoding, number line estimation, and (non)symbolic number magnitude comparison tasks. Discriminant analysis will be employed to assess whether basic number processing tasks can differentiate between a healthy control group and both PD groups. All participants will be subjected to a comprehensive numerical and a neuropsychological test battery, as well as sociodemographic and clinical measures. Study results will give the first broad insight into the extent of basic numerical deficits in different PD patient groups and will help us to understand the underlying mechanisms of the numerical deficits faced by PD patients in daily life.

KEYWORDS
basic number processing, mild cognitive impairment, numerical cognition, Parkinson's disease, triple code model
Modern societies currently face the challenges of demographic changes, with age-associated diseases bringing medical, financial, and psychological burdens to patients, caregivers, and the healthcare system. In the elderly population above 60 years of age, the overall prevalence of mild cognitive impairment ranges between 12% and 18% (Petersen, 2016), whereas dementia affects 5% to 7% (Prince et al., 2013). Severe cognitive impairment and dementia are associated with a loss of independent functioning and social and legal competence (Sherod et al., 2009). Therefore, supporting active participation in society and functional independence of the elderly people (with cognitive deficits) seems to be one of the most crucial challenges.

Parkinson’s disease (PD) is characterized by apparent motor deficits like bradykinesia, rigidity, and resting tremor (Postuma et al., 2015). The presence of nonmotor symptoms is common in PD patients and has a substantial impact on their quality of life (Prakash et al., 2016). The development of Parkinson’s disease dementia (PDD), especially in more advanced PD disease stages, leads to an increased risk for nursing home placement and mortality (Aarsland et al., 2000; Bjørnestad et al., 2017). The presence of mild cognitive impairment in PD patients (PD-MCI) is associated with a shorter window of time before PDD onset and is therefore considered as one of the most important risk factors for dementia (Aarsland et al., 2017). The onset of cognitive dysfunction varies over the disease course, and deficits can sometimes already be identified at the time of PD diagnosis using sensitive instruments (Kehagia et al., 2010; Martinez-Horta & Kulisevsky, 2019).

Societal burdens caused by PD on a regional to global scale clearly show the need to develop diagnostic instruments for the disease-immanent cognitive deficits to anticipate the burden on patients and caregivers (Dorsey et al., 2018). Considering the most recent knowledge about PD neuropathology and symptoms, different cognitive biomarkers might serve this purpose (Roheger et al., 2018).

1.1 | Parkinson’s disease—Epidemiology and clinical symptoms

Parkinson’s disease is the second most common neurodegenerative disease worldwide with over 6 million affected patients globally (Dorsey et al., 2018). Prevalence differs depending on sex and age with a higher incidence in men and the elderly people (Hirsch et al., 2016). In terms of cognitive symptoms, the prevalence of PDD in the United States is predicted to at least double between 2015 and 2060 (Savica et al., 2018). All of these epidemiological estimates show that PD already signifies a huge concern for the public health system and will become even more crucial in future due to demographic changes (Dorsey et al., 2018; Kowal et al., 2013).

The apparent cardinal symptoms of PD are hypokinesia and bradykinesia alongside at least rigidity and/or resting tremor (Postuma et al., 2015). Nonmotor symptoms can include dysautonomia (e.g., incontinence, sleep disorders) or be psychiatric in nature (e.g., depression; Jankovic, 2008). Furthermore, the cognitive symptoms discovered thus far imply deficits in the domains of executive functions, working memory and attention, memory, language, and visuospatial functions (Aarsland et al., 2017; Litvan et al., 2012).

Further differentiation of cognitive deficits in PD can be pursued on a continuum ranging from normal cognition (PD-NC) over the initial stage of clinically significant cognitive disorders (PD-MCI) to PDD, implying a therapeutic indication. The diagnosis of PD-MCI is based on impairments in global cognition or in at least one of the domains of executive functions, working memory and attention, memory, language, and visuospatial functions (Emre et al., 2007; Litvan et al., 2012). Deficits in PDD additionally affect the activities of daily living (ADL) such as dressing, housekeeping, and financial management (Christ et al., 2013).

Cognitive disorders in PD patients are characterized by marked heterogeneity, with 30% to 60% moving from PD-NC to PD-MCI or PD-MCI to PDD within 3–4 years (Janvin et al., 2006; Lawson et al., 2017). The development of PD-MCI can arise from frontostriatal dysfunction (triggered by dopaminergic deficits) and posterior cortical impairments (caused by Lewy bodies and nondopaminergic deficits) with the latter being most predictive of PDD (Barker & Williams-Gray, 2014). Whether this posterior cortical profile describes one homogenous etiology is still being investigated (Monchi et al., 2016). The cognition of PD patients is further impacted by factors such as education, aging, comorbidities, and gender (Lin et al., 2018). According to recent findings, one of the most important risk factors for PDD—besides male sex, old age, and severity of motor symptoms—is the occurrence of PD-MCI (Aarsland et al., 2017). Therefore, the development of sensitive instruments for the identification of PD-MCI patients with a high risk of PDD progression is necessary to initiate timely individual and societal compensatory and supportive strategies (Martínez-Horta & Kulisevsky, 2019). Surprisingly, the domain of numerical cognition...
has been neglected despite the apparent deficits of PD patients observed in clinical practice.

At the PD-MCI stage, cognitive deficits are mild enough to be addressed by therapeutic interventions (Leung et al., 2015). The aim of cognitive training is to improve daily functioning and increase self-sufficiency. Within this framework, a very important domain necessary for daily functioning is numerical literacy, including financial capabilities, time management, and household planning. The management of complex numerical and arithmetic processes requires the integration of different basic number processing functions. If these functions are impaired, deficits need to be considered in therapy, remediation, and intervention. However, the lack of research in this field does not allow for any conclusions to be drawn. For this reason, the current study will systematically investigate deficits in the basic number processing abilities of PD patients based on established neurocognitive theories.

### 1.2 Theories of basic number processing

Possible deficits in the number processing abilities of PD patients can be characterized by looking at the underlying functions of non-pathological numerical cognition. Neurocognitive models of number processing (i.e., the triple code model and its extensions; Dehaene & Cohen, 1995; Dehaene et al., 2003; Klein et al., 2016) consider number magnitude as the core representation of numerosity, which can be further differentiated. The approximate number system (ANS) is supposed to underlie basic numerical operations, whereas magnitude is also represented symbolically for more accurate processes (Cantlon et al., 2009; Dehaene, 2001, 2009). Whether the same number module underlies these two (non)symbolic number magnitude representations is controversial (Fias et al., 2003; Holloway et al., 2010; Lyons et al., 2012; Matejko & Ansari, 2016). The acuity of the ANS can be evaluated using nonsymbolic number magnitude comparisons. Both symbolic and nonsymbolic number representations can be addressed by studying distance effects in (non)symbolic number magnitude comparison tasks.

Furthermore, there are visuospatial magnitude representations (Dehaene et al., 2003; Nuerk et al., 2011). These can explain the so-called spatial-numerical associations, where different numerical attributes are associated with space (Cipora et al., 2018). Number line estimation can be used to assess this visuospatial representation of magnitude.

An additional prerequisite for the processing of multi-digit numbers is place \( \times \) value integration. For the appropriate magnitude judgment as well as the application of mathematical operations on multi-digit numbers, distinct places and corresponding values need to be identified, manipulated, and/or integrated (Nuerk et al., 2015). Place \( \times \) value integration is revealed by the occurrence of specific errors in transcoding tasks as well as the unit-decade compatibility effect in symbolic number magnitude comparisons of two-digit Arabic numbers. To investigate all of these representations of number processing, an effect-based approach is preferred over a task-based approach and will be used in the current study (Nuerk et al., 2011).

For basic number processing to be performed successfully, both domain-specific numerical and domain-general functions are necessary (covered by the frontoparietal network of number processing; Klein et al., 2016). On the one hand, deficits in basic number processing can be primarily based on domain-specific numerical functions such as number magnitude representation, place-value integration, or knowledge of mathematical procedures. On the other hand, secondary deficits in basic number processing can also result that are triggered by deficits in domain-general functions such as attention, working memory, or executive functions, as these functions are a prerequisite for successful number processing (Knops et al., 2017). Although there is evidence for domain-general deficits in PD patients (Litvan et al., 2012), the effect of PD-characteristic neurodegeneration on domain-specific numerical cognition will be investigated here for the first time.

### 1.3 Basic number processing and Parkinson's disease

Despite the lack of research on basic number processing of PD patients, hypotheses about lost or preserved functions can be generated based on nonpathological aging effects/deficits in other neurodegenerative diseases such as Alzheimer's disease (AD), and the neuroanatomical and functional overlap between basic number processing and PD neuropathology.

Considering human ontogenesis, number processing changes with age. Domain-general factors such as attention, working memory, processing speed, and executive functions play a role in healthy aging processes concerning number processing (Hinault & Lemaire, 2016; Li et al., 2004; Schretlen et al., 2000). Numerical deficits associated with aging were found in the precision of the ANS (Halberda et al., 2012) and in financial abilities (Finke et al., 2017). This illustrates the implications of domain-specific and domain-general deficits in healthy numerical aging processes and is the reason why the current study uses a healthy elderly control group.

Research on AD has already identified several pathophysiological mechanisms and predictive biomarkers. The diversity of the deficits experienced by AD patients indicates that there may be heterogeneous profiles in the domain of number processing and the overlap with PDD neuropathology is only partial (e.g., the cholinergic system; Bohnen et al., 2003). Therefore, conclusions from these studies are informative, but limited, for PD (Hugo & Ganguli, 2014).

Deficits in the number processing abilities of AD patients are unique, with the occurrence of specific intrusion errors in transcoding. For this reason, transcoding tasks are suggested for use as part of the AD diagnostic procedure (Kalbe, 1999; Rosselli et al., 1998). AD patients also show compensatory brain activation mechanisms shifting to premotor areas due to parietal dysfunction (Ouchi et al., 2004). This shift in neuronal activation might differ for PD, as neurodegeneration in PD also involves the premotor cortex (Lindenbach & Bishop, 2013).
Numerical activities of daily living (NADL) are crucial for the economic and legal literacy of patients affected by neurodegeneration (Sherod et al., 2009). As the ability of AD patients to complete the associated tasks such as financial management, public transport, or medication management deteriorates early in the disease process (Martini et al., 2003; Sherod et al., 2009), the systematic investigation of deficits in number processing in the context of neurocognitive diagnoses is needed for medicolegal evaluations (Simpson, 2014). In most cases when ADL deficits become apparent in a PD patient, the individual has progressed from a state of PD-MCI to PDD (Emre et al., 2007). Because of their negative impact on the quality of life, depression, mortality, and caregiver burden, ADL deficits in PD patients need to be identified as early as possible (Lerol et al., 2012). In order to fully understand the genesis of NADL deficits, basic number processing needs to be investigated thoroughly as it represents an earlier stage of numerical impairment.

Kalbe (1999) showed that PDD patients have impairments in counting, transcribing, and number line estimation. It can only be inferred that some aspects of basic number processing are impaired in PD at the advanced stage of dementia; however, whether deficits are present at earlier disease stages remains unclear. The task-based approach applied by Kalbe (1999) also does not allow for conclusions regarding representations of basic number processing which could be achieved with an effect-sized approach. Additionally, Kalbe’s study was conducted with small sample sizes, decreasing generalizability to the population level. It also focused on inferences regarding AD, despite the PDD group showing worse performance than the control, AD, and vascular dementia groups in seven out of 14 numerical tests. The comparison of the performance of nondemented PD patients with healthy elderly individuals on complex arithmetic tasks showed the presence of several deficits (Liozidou et al., 2012; Martin et al., 2013; Zamarian et al., 2006). These findings indicate the need to systematically investigate which basic numerical representations might be the underlying cause of complex arithmetic deficits and how the course of degeneration affects both PD-NC and PD-MCI patients.

Due to the lack of information on basic number processing in nondemented PD patients, hypotheses to be investigated in the current study can be generated based on the neuroanatomical and functional overlap between basic number processing and the progression of PD neuropathology. The neuropathology of PD is characterized by the degeneration of the basal ganglia, which are responsible for dopamine production and are active in neuronal loops of motor control and number processing (Moeller et al., 2015; Moustafa et al., 2014). Additionally, degeneration occurs in the cerebellar and additional subcortical and cortical brain structures, depending on the stage of disease progression (categorized using Braak stages ranging from one to six; Braak et al., 2003; Caligiore et al., 2016). On the molecular level, an imbalance of the dopaminergic, cholinergic, noradrenergic, and serotonergic systems is accompanied by a characteristic abnormal cortical and subcortical accumulation of protein, the so-called Lewy bodies (Braak et al., 2003). The variety of affected neuronal structures and functions explains diagnostic challenges and the complexity of PD symptoms.

In the course of PD progression, brain areas, which are necessary for basic number processing, become affected sequentially. Several of these areas are involved in domain-specific numerical function so their degeneration implies that patients may exhibit the processes of primary dyscalculia. The reduction of acetylcholine production in the nucleus basalis magnocellularis (beginning in Braak stage 3) is associated with a weakening of projections to the neocortex, such as parietal and temporal areas (beginning in Braak stage 5; Jellinger, 2018). This might affect the responsibility of the intraparietal sulcus for (non)symbolic number magnitude representation and place-value processing as well as superior parietal regions needed for visuospatial representations on a mental number line (Jellinger, 2018; Moeller et al., 2015; Winter et al., 2015). Furthermore, visual number form might be affected and is represented in the temporal cortex (Arsalidou & Taylor, 2011; Koob et al., 2014). Inflammation in the angular gyrus and Lewy body-induced degeneration of the hippocampus and posterior cingulate gyrus might affect retrieval from verbal memory and the verbal representation of numerosity (Jellinger, 2018; López Gonzalez et al., 2016; Uribe et al., 2018). Lewy Bodies in frontal, cingulate, and the temporal cortex emerging in Braak stage 5 further impact visuospatial functions and number processing (Arsalidou & Taylor, 2011; Collerton et al., 2003; Klein et al., 2016). In Braak stage 6, lesions in first-order sensory association areas of the neocortex and premotor areas occur, occasionally accompanied by mild changes in primary sensory areas and the primary motor field (Braak et al., 2003). The decrease in dopamine levels further impacts intracortical inhibition in the motor cortex, worsening time estimation (Lindenbach & Bishop, 2013). These processes can be linked to deficits in the embodiment of number processing and phonological processing in premotor and supplementary motor areas (Winter et al., 2015). The degeneration of domain-general functions necessary for basic number processing can lead to secondary dyscalculia in PD patients, with degeneration occurring in early disease stages. Dopamine production beginning to decrease in the striatum in Braak stage 3 later results in reduced projections to the frontal lobe and consequently deficits in verbal fluency, processing speed, working memory, memory retrieval, attention, and executive functions (Braak et al., 2003; Dirnberger et al., 2005; Martínez-Horta & Kulisevsky, 2019; Rinne et al., 2000).

### 1.4 Objectives of the current study

The current evidence regarding numerical deficits in PD is scarce and the differences between PD-NC and PD-MCI have not been investigated at all. Therefore, the aim of the current study was to identify deficits in the basic number processing abilities of PD patients with and without mild cognitive impairment as compared to a healthy control (HC) group. To better understand the underlying mechanisms of behavioral deficits, the influence of both domain-general and domain-specific functions will be addressed. The final aim was to investigate whether performance on basic numerical tasks can be used to differentiate between the different cognitive
status of HC, PD-NC, and PD-MCI and thus be applied in the diagnostic assessment of PD-immanent cognitive disorders.

As level I-PD-MCI based on cognitive screening (level I-PD-MCI) patients are situated at a point of considerable cognitive impairment, they are expected to exhibit worse performance than HC and PD-NC groups. Whether performance of the PD-NC group is closer to the HC or the level I-PD-MCI group will be examined.

The above considerations resulted in the following research questions:

1. What effects and representations of basic number processing are impaired in PD patients? (H1: Basic numerical deficits in PD)
2. Can basic number processing performance of PD patients be explained by domain-general functions? (H2: Influence of domain-general functions on basic number processing in PD)
3. Can basic number processing performance be used to discriminate between HC, PD-NC and level I-PD-MCI patients? (H3: Discrimination between the cognitive statuses of PD by the tests of basic numerical cognition)

Basic numerical processes will be investigated with the help of (a) transcoding and (b) number line estimation tasks, and (c) nonsymbolic and (d) symbolic number magnitude comparisons.

2 | METHODS

2.1 | Statistical power analysis and sample size estimation

The sample will be composed of the three groups HC, PD-NC, and level I-PD-MCI. Required sample sizes for each hypothesis were computed based on empirical estimates using G*Power (Faul et al., 2007, 2009). Effect sizes were calculated and transformed using Psychometrica (Lenhard & Lenhard, 2016).

Systematic studies on the impairment of basic number processing in PD, which can be used for effect size calculations, are scarce. Kalbe (1999) reports differences with large effect sizes between a HC and a PDD group in transcoding \((d = 1.755)\), number line estimation \((d = 0.868)\), and magnitude comparison \((d = 1.051)\). The available effect sizes are expected to be overestimates due to publication bias and the comparison with a cognitively more deteriorated PD patient sample than the ones to be tested in the current study. Therefore, the smallest effect size was lowered to the conventional effect size for large effects of \(d = 0.8\) (Cohen, 1992) for separate power analyses estimating the required sample size for the three research questions. Consequently, parameters for the effect size calculations were set to \(d = 0.8, \alpha = 0.05, \) and \(\text{power} = 0.9\). In the case of participant exclusion, new participants will be recruited for substitution. In case of early ceasing of the testing due to attrition, new participants will be recruited, and data of the dropped out patient will only be included in analyses when the patient gives informed consent and has already been assigned to a cognitive group based on the MoCA classification. As performance of PD patients in cognitive tests can be severely affected by other clinical confounders (e.g., age, gender, education, UPDRS-III and/or Hoehn and Yahr staging, disease duration, depression), a maximum of six covariates will be included in the analyses.

2.1.1 | Basic numerical deficits in PD (H1)

The ANCOVA sample size was calculated for a comparison between three groups with an effect size of \(f = 0.4\) (equivalent to \(d = 0.8\)) for a maximum of three independent variables assuming a maximum of six clinical confounders. This results in a minimum sample size of \(n = 111\) participants.

2.1.2 | Influence of domain-general functions on basic number processing in PD (H2)

The question of whether performance on numerical tasks can be explained by performance on domain-general tests will be addressed with further ANCOVAs. ANCOVA sample size was calculated for a comparison between three groups with an effect size of \(f = 0.4\) (equivalent to \(d = 0.8\)) for a maximum of three independent variables assuming a maximum of six clinical and one cognitive (i.e., seven in total) covariates. This results in a minimum sample size of \(n = 117\) participants.

2.1.3 | Discrimination between cognitive statuses of PD by the tests of basic numerical cognition (H3)

The last diagnostic discriminant analysis between HC, PD-NC, and level I-PD-MCI will be carried out using a logistic regression. The logistic regression sample size was calculated for an effect size of \(OR = 4.27\) (equivalent to \(d = 0.8\)) for a maximum of 17 predictors assuming normal distributions. This results in a minimum sample size of \(n = 39\) participants.

For testing all three hypotheses, the same subjects will be used with a total sample of at least \(n = 117\) valid data sets (as required for the most complex analysis for H2). Number of patients per group should be approximately \(n = 39\) participants, but this might vary slightly due to the recruitment process and the outcome of the cognitive diagnosis based on the neuropsychological test data. We were not able to estimate effect sizes for comparison between the two PD groups, as there are no data published yet on this topic.

2.2 | Participants

This study has been approved by the ethics committee of the University of Tübingen’s medical faculty (161/2020BO2) and has been registered at the Deutsches Register für Klinische Studien
(DRKS-ID: DRKS00021091) and with the World Health Organization (Universal Trial Number: U1111-1257-2901). Patients will be recruited via the PD outpatient clinic and in collaboration with rehabilitation facilities specialized in PD. In addition, previously studied PD patients who gave their consent (Ethical vote: 199/2011BO1) to be contacted in case of further potential study participation will be contacted. Patients’ caregivers will be recruited as HCs according to defined inclusion and exclusion criteria. Furthermore, pensioners’ initiatives will be contacted for control group recruitment and the study will be advertised using the university mailing system. All participants will receive monetary compensation. In the recruitment process, participants will be matched on the group level according to the following criteria: age (M±5 years) and gender (max. 65% male). This will result in sociodemographic and clinical group means that will be as similar as possible. Disease duration cannot be corrected for in this way because cognitive status is confounded with disease duration in PD-MCI patients, who also have a longer PD history as indicated by PD-specific disease progression (Lin et al., 2018). As PD medication can heavily influence experimental performance, all patients will be tested in their “on-state” and will be allowed to take their regular medication during the session if necessary.

Inclusion and exclusion criteria are listed separately for all three groups in Table 1. Idiopathic PD diagnosis according to the United Kingdom Brain Bank criteria (Hughes et al., 1992) needs to be confirmed by a movement disorder specialist in the outpatient clinic of the University Hospital of Tübingen. Verification of the diagnosis is required before inclusion in the current study and needs to be documented in a patient’s record. The assessment of the UPDRS-III motor examination will be conducted based on a training in the PD outpatient clinic. The HC group will be recruited as similarly as possible to the two PD groups in regard to age, education, and gender. Additionally, a spouse, child, other relative, or friend of legal age of all PD patients who agreed to report on the presence and severity of PD-related ADL problems (see section caregiver assessment) will be asked for an assessment of the patient.

The cognitive diagnosis leading to assignment to the PD-NC group, the level I-PD-MCI group, or exclusion in the case of PDD will be carried out according to the level I MDS Task Force criteria for PD-MCI and PDD (Emre et al., 2007; Litvan et al., 2012). This implies a cutoff score of 26 on the screening tool Montréal Cognitive Assessment (MoCA; Nasreddine et al., 2005) to differentiate between PD-NC and level I-PD-MCI and absence of significant ADL impairment impacting everyday life to rule out PDD. The investigator HL will verify the diagnosis of each participant based on the MoCA cutoff criteria with the experienced neuropsychologist ILS. The impairment of ADL is operationalized with an FAQ cutoff score of ≥3 (see Becker et al., 2021). For the cognitive diagnosis, patients and caregivers will be asked whether they have noticed progressive deterioration of the patient’s cognitive state. The same MoCA cutoff score will be used for the exclusion of individuals with cognitive

| (Sub)group | Inclusion criteria | Exclusion criteria |
|------------|-------------------|-------------------|
| General    | ≥60 years of age  | Further neurological diseases impacting the central nervous system except for damage to intervertebral discs |
|            | (Corrected) hearing and vision | (History of) substance abuse except for nicotine |
|            | Fluent German skills and German schooling | BDI-II-Score ≥ 20 indicating major depression |
|            | Informed consent, voluntary participation | Delirium or acute psychosis |
|            | Stable health status (i.e., able to conduct the entire testing, comorbidities may be present but do not affect test performance) | Acute psychiatric diseases |
|            | Permission to ask a caregiver to verify patients rating on the occurrence of ADL problems | 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 2 second degree relatives) |
|            |                   | Currently undergoing chemotherapy |
| HC         | Normal cognition (MoCA ≥ 26) | Severe impulse control disorder, dopamine dysregulation syndrome interacting with the patient’s everyday life |
| PD         | Diagnosed idiopathic Parkinson’s syndrome | Diagnosis of cognitive impairment (level I-PD-MCI based on cognitive screening or PDD; Emre et al., 2007; Litvan et al., 2012) |
| PD-NC      | Normal cognition according to MDS Task Force level I diagnosis guidelines (e.g., MoCA ≥ 26) | Diagnosis of PDD (Emre et al., 2007) |
| Level I-PD-MCI based on cognitive screening | Mild cognitive impairment according to MDS Task Force level I diagnosis guidelines and diagnostic cutoff point based on Hoops et al. (2009; 26 > MoCA >18) |  |
impairment from the HC group. The MoCA (Nasreddine et al., 2005; Thomann et al., 2018) is a short screening instrument for global cognitive functions, evaluating short-term memory, visuospatial and executive functions, attention, language, and orientation to time and space. It has a maximum score of 30 and uses education status correction for patients with 12 years of education or less.

2.3 | Materials

Our materials and programmed experiments will be openly available via the Open Science Framework (https://osf.io/ap6je/). If the participants gave their consent for study inclusion, inclusion and exclusion criteria will be checked during a clinical visit or via a telephone call using a semi-standardized questionnaire.

2.3.1 | Introductory interview

An introductory interview will be used at the beginning of experimental sessions to acquire information on sociodemographic (age, gender, years of education, handedness, and mother tongue) and clinical variables (diagnosis, age at disease onset, medication for depression and cognition). Furthermore, levodopa equivalence dose (LEDD; DGN, 2016; Tomlinson et al., 2010) will be calculated based on current dopaminergic medication and time since last dopaminergic medication intake will be recorded. Regarding experience with mathematics, participants will be asked for 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 will be asked whether they have noticed a progressive deterioration of their cognition.

2.3.2 | Motor performance

Performance concerning PD motor symptoms will be evaluated with the sum score of the MDS revision of the Unified Parkinson’s Disease Rating Scale Part III and IV and the items Tremor, Walking and balance, and Freezing from Part II (UPDRS-II, III & IV; Goetz et al., 2008) and Hoehn and Yahr staging (Hoehn & Yahr, 1967). The motor examination of UPDRS-III covers 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. Evaluation follows a scale ranging from 0 = normal to 4 = severe, with high values indicating more pronounced motor impairment. A total score maximum of 132 based on 18 items (with several items relating to multiple extremities resulting in 33 scores) will be calculated and included in the analysis. The subtest UPDRS-IV will be carried out to assess dyskinesia, motor fluctuations, and dystonia based on six items with a score ranging from 0 to 24 points. The Hoehn and 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) will be assessed as an additional measure for PD severity. Along with performance on the UPDRS-III, the items from UPDRS-II are needed to classify patients as being of the tremor or postural instability and gait disorder motor type (Stebbins et al., 2013).

2.3.3 | Neuropsychological test battery

Performance in the cognitive domains of executive functions, working memory and attention, verbal memory, language, and visuospatial functions will be assessed with at least one test per domain. For each test, raw scores will be included as covariate measures.

For executive functions, inhibition will be assessed with the subtest Go/No Go “2 out of 5” in the test battery of attention (TAP = Testbatterie zur Aufmerksamkeitsprüfung; Zimmermann & Fimm, 2017). Participants are required to discriminate five types of stimuli and to react to only two of them. Performance on this test will be measured by the number of errors.

Working memory abilities will be assessed in a letter span forward and backward task (as in Soltanlou et al., 2015). Participants will have to listen to letter strings of increasing length. In the first part, they will have to reproduce them in the same order, and in the second part, the letters have to be reproduced in reverse order. The relevant outcome variable will be the maximum length of reproduced letter strings backwards. Visuospatial working memory will be assessed with the Corsi block-tapping test, requiring to mimic a sequence of blocks tapped on by the experimenter in the correct forward and backward order (Corsi, 1972; Kessels et al., 2000). Sequences increase starting from two blocks with a maximum of nine, when both items of the same length are correctly reproduced, and the outcome measure is the longest Corsi backward span correctly reproduced. Attention will be measured with the TAP subtest Alertness (Zimmermann & Fimm, 2017). The test requires a simple reaction to a visually presented stimulus with or without prior notice via an auditory cue. The median RT in conditions without an auditory cue (as a measure of intrinsic arousal) will be used as the outcome variable.

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

The German version of the WAIS-IV subtest Similarities (Petermann, 2012) measures language operationalized as the...
conceptual understanding of two words by indicating what they have in common. The total number of correctly solved items will be used as a measure of word knowledge.

The Benton Line Orientation Test (Benton et al., 1978) is a measure of visuospatial function where two lines of a certain angle and position are presented on a sheet of paper. Patients are supposed to compare these two lines, with 11 lines being arranged in a star-shaped fashion around a center. Identifying the two lines out of the displayed 11 lines correctly is scored as 1 point and a maximum sum score of 15 can be achieved.

### 2.3.4 Numerical tasks

Basic number processing will be assessed with computerized procedures and a paper-and-pencil test. The tasks are typical tasks used to assess basic number processing and the related numerical representations. The current study design requires a detailed assessment instead of an overall screening, and the control group will be used to compensate the lack of standardized norms. The required response mode is adjusted to the PD-immanent motor impairments in as much as cards with single-digit numbers will be used for transcoding to avoid handwritten responses. Additionally, responses in the reaction time tasks are given with easier-to-handle TAP keys (magnitude comparisons) and the mouse (number line estimation). There will be no time restrictions in any of the tasks, but testing will be stopped if a participant cannot answer the first five experimental trials of the current task.

The **transcoding** task consists of transforming auditory number words into written Arabic digits (16 items) as well as written Arabic digits into written number words (16 items) in a paper-and-pencil format, with the experimenter presenting items and coding responses manually. The experimental items will be preceded by a practice block of three trials each. The outcome measures will be overall accuracy and number of errors per category (error categories similar to Kalbe, 1999; Zuber et al., 2009). Lexical errors address the meaning of the digit (i.e., 24 → 25, 90 → 91, 90 → 19). Syntactic errors imply the composition of multi-digit numbers (i.e., additive composition: 360 → 30,060, multiplicative composition: 400 → 4,100, and inversion ignored: 182 → 128). Dementia-specific errors are intrusion errors with switches of the response modality (i.e., 2007 → 2thousand7) or intrusions from the preceding item. Orthographic errors are categorized as education- and disease-dependent errors.

The accuracy of a range-dependent number-to-space mapping will be assessed in a computerized number line estimation task. Participants will have to locate 20 given numbers per number line (range: small 0–100, large 0–1,000,000) with the arrow keys of a standard QWERTZ keyboard. For each range, numbers will be equally distributed over the entire number line and the percent absolute error per range will be calculated as the measure of (in)accuracy.

The ANS will be assessed with a nonsymbolic number magnitude comparison task. Two groups of dots will be presented simultaneously next to each other for 200 ms. This presentation phase will be followed by the participant’s response with a left- or right-hand key press to indicate which group of dots is larger. After an interstimulus interval of 200 ms, the next trial will be presented. Proportions between the two groups of dots will comprise four different ratios (1:2, 3:4, 5:6, and 7:8) with 48 trials each and a minimum group size of four. Furthermore, stimuli will be manipulated based on the visual attributes’ total occupied area, individual dot size, convex hull area, and mean occupancy (i.e., the inverse of density) in a balanced design within each ratio (see Guillaume et al., 2020). Performance will be evaluated with average reaction time per ratio and the Weber fraction as a measure of accuracy with small fractions indicating high accuracy (as in Dietrich et al., 2015).

A symbolic number magnitude comparison task will be used to measure the exact magnitude representation. Participants have to respond with the TAP keys positioned above each other to indicate which of two numbers (presented vertically aligned, one above the other on a screen) is larger. The duration of stimulus presentation will be until the participant presses the response key. Stimuli will be manipulated by unit-decade compatibility (comparison of units and decades separately would have the same result as the whole number) and distance (large ≥ 40; as in Moeller et al., 2012). The number pairs will consist of 60 unit-decade compatible and 60 unit-decade incompatible two-digit Arabic numbers as well as 120 filler items (numbers within the same decade). The outcome measures will be the average reaction times for compatible versus incompatible trials and small versus large distances. This allows for the assessment of the unit-decade compatibility effect (faster reaction to compatible trials) and the distance effect (faster reaction to large distances; Nuerk et al., 2001). Furthermore, overall accuracy will be calculated.

### 2.3.5 Clinical questionnaires for nonmotor assessment, ADL, and quality of life

In addition to the introductory interview, further clinical variables will be assessed with self-report questionnaires. The Nonmotor Symptoms Questionnaire for Parkinson’s disease (NMSQuest; Chaudhuri et al., 2006) is a 30-item screening questionnaire with a categorical answer format of “yes,” “no,” and “don’t know” based on the occurrence of nonmotor symptoms in the last month. It allows quantification of sleep disorders, and neuropsychiatric, autonomic, gastrointestinal, sensory, and other symptoms, with the help of a sum score.

Participants’ depressive symptoms will be rated with the Beck Depression Inventory (BDI-II; Hautzinger et al., 2006) with a cutoff of 20 indicating signs for major depression. The inventory consists of 21 items to be rated for the intensity of occurrence of the symptom on a scale ranging from 0 to 3 (maximum sum score of 63). This questionnaire reflects the patient’s feelings during the last 2 weeks.

The Functional Activities Questionnaire (FAQ; Pfeffer et al., 1982) is used to measure current social function in older adults. Participants will have to rate their level of performance (ranging from 0 = normal to 3 = dependent) in 10 different daily life activities, resulting in a sum score characterizing social function.

Health-related quality of life will be 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, will be coded on a scale ranging from 0 (= perfect health) to 100 (= worst health).

2.3.6 | Caregiver assessment

For a broader evaluation of participant quality of life, the FAQ will also be filled out by a caregiver. Furthermore, caregivers will be asked for their relationship with the participant, their demographic data (i.e., age and gender), and the frequency, as well as the intensity, of contact with the participant. For the cognitive diagnosis, caregivers will be asked whether they have noticed a progressive deterioration of the patient’s cognitive state.

2.4 | Procedure

Participants will give written informed consent. In the second step, participant eligibility will be assessed according to the defined inclusion and exclusion criteria. PD patients will be assigned to PD-NC or level I-PD-MCI groups according to their cognitive performance. As the current study belongs to a larger project, the testing will be conducted together with further measures which are not included in the current publication. Participants will attend two sessions of 1.5 to 2 hr each. In order to handle patient attrition, there will be breaks within each session when patients need them. In the first experimental session, the sociodemographic questionnaire, and the numerical tasks transcoding, number line estimation, and nonsymbolic and symbolic number magnitude comparison will be carried out in this order. Afterward, participants and caregivers will complete the clinical scales and questionnaires which may also be filled out at home between the first and second session. The second session consists of the MoCA, clinical variables, motor assessment, and neuropsychological test battery in this order. The two sessions may be scheduled 3 weeks apart at maximum.

3 | DATA TREATMENT AND PROPOSED ANALYSIS PIPELINE

Data analyses will be run with R version 3.4.1 (R Core Team, 2014), and JASP version 0.11.1.0 (JASP Team, 2018). Data will be managed using REDCap electronic data capture tools (Harris et al., 2009). Anonymized data and analysis scripts will be freely available on the Open Science Framework (https://osf.io/ap6je/). For all inferential statistics, an α-level of 0.05 will be assumed and effect sizes will be estimated.

3.1 | Data preprocessing

3.1.1 | Exclusions

Participants with missing data will be excluded in a case-wise manner for the analyses including the respective measure. For forced choice format tasks (nonsymbolic and symbolic number magnitude comparison, TAP tasks, and Benton Line Orientation Test), participants need to achieve an accuracy (ACC) of at least 75% to be included in the analysis of the respective measure. In the number line estimation task, understanding of task instruction will be checked for by monotony in responses, with participants being excluded if they do not answer overall in a format of numbers ascending from left to right (indicated by a minimum positive space-magnitude correlation of 0.3). Understanding of task instruction in transcoding will be assessed with a minimum accuracy of 50% in single-digit trials. Furthermore, participants will be excluded from the respective analysis if their performance exceeds 3 SD below the group M.

3.1.2 | Reaction times

Data trimming for reaction times (RTs) will follow Baayen and Milin (2010). After the exclusion of incorrectly solved trials, the RT distribution will be inspected with per-subject quantile–quantile plots. This method is pursued to identify the theoretical model for data transformation in order to achieve a normal distribution. The most appropriate distribution will be chosen based on the best model fit and used to transform the data. The next step considers two levels of outlier exclusion. Physically impossible RTs will be considered to be anticipations faster than 200 ms and be excluded. Then, outliers will be excluded based on model criticism. This will be achieved with Shapiro tests for normality. Those data points with absolute standardized residuals exceeding 3 will be removed. The last step consists of corrections for temporal dependencies. First, autocorrelation functions by participant will be carried out. Afterward, a regression model will be fitted to responses with a log-transformation for latencies including the covariates trial number and preceding RT.

3.1.3 | Accuracy

Depending on the best model fit for the empirical distribution of accuracies, these data will either be arcsine- or logit-transformed.

3.1.4 | Assumption check

After these separate steps of data preprocessing, specific assumptions for statistical hypothesis testing will be checked. In the case of a violation of linear model assumptions for the ANCOVA, a robust ANCOVA will be conducted. Predictors of the logistic regression will be checked for collinearities based on a variance inflation factor below 10.

3.2 | Group characteristics

Sociodemographic, clinical, and cognitive variables will be compared between the three groups in order to identify possible confounding...
differences between them. The categorical variables gender, motor type, and Hoehn and Yahr staging will be characterized as total amount per category and corresponding percentage and compared with chi-squared tests between HC and PD-NC and between PD-NC and level I-PD-MCI groups. Continuous variables will be described with M (SD) and compared between HC and PD-NC groups and between PD-NC and level I-PD-MCI groups with an independent samples t test. The frequentist group comparisons will be supported with the calculation of Bayes factors. These variables are sociodemographic (age and education years) and clinical (disease duration, age at onset, LEDD, and intake of antidepressants) variables. Furthermore, motor (UPDRS-III & IV, motor type) and cognitive function will be compared. The last group comparison addresses sum scores of completed questionnaires (nonmotor symptoms: NMSQuest, depression: BDI-II, ADL self-report and caregiver report: FAQ, health-related quality of life: PDQ-39). At this point, we will report the number of HC, PD-NC, and level I-PD-MCI participants being impaired in any of the cognitive measures.

3.3 | Hypothesis testing

The three groups will first be compared regarding sociodemographic and clinical variables. We assume age, gender, education, UPDRS-III (with motor type), and/or Hoehn and Yahr staging, disease duration, and depression could significantly differ between groups. In this case, we will include a maximum number of six clinical covariates in all models. The maximum number of covariates will be set to seven for the analysis of H2, allowing the additional integration of the cognitive variable with the largest association (minimum correlation of 0.3) for each numerical task separately. We expect to find an ordinal trend for the HC group outperforming both PD-NC and level I-PD-MCI groups, and the PD-NC group outperforming the level I-PD-MCI group. For all numerical tasks concerning hypotheses 1 and 2, the ordinal trend of performance decreasing from HC over PD-NC to level I-PD-MCI groups will be tested with pairwise ANCOVAs that include the covariates of the respective analysis.

3.3.1 | Basic numerical deficits in PD (H1)

The transcoding task will be analyzed using an ANCOVA with the factor group (HC, PD-NC, or level I-PD-MCI) and the dependent variable ACC. Error categories (lexical errors, syntactic errors, dementia-specific errors, and education- and disease-dependent errors; no rest category) will be compared between groups (HC vs. PD-NC, PD-NC vs. level I-PD-MCI, and HC vs. level I-PD-MCI) with chi-squared tests. Performance in number line estimation will be assessed by first calculating individual regression slopes per range (small vs. large) using percent absolute error. Then, a 3 (group: HC, PD-NC, and level I-PD-MCI) × 2 (range: small vs. large) ANCOVA will be conducted with the slope parameter as the dependent variable. Nonsymbolic number magnitude comparison will be evaluated with two separate ANCOVAs of the factor group (HC, PD-NC, and level I-PD-MCI) on the dependent variables Weber fraction and overall ACC. The analysis of symbolic number magnitude comparison will be conducted both for the dependent variables RT and ACC by separate 3 (group: HC, PD-NC, and level I-PD-MCI) × 2 (distance: small vs. large) × 2 (unit-decade compatibility: compatible vs. incompatible) ANCOVAs. Following this frequentist approach, Bayesian ANCOVAs will be calculated to allow for further conclusions on the presence or absence of effects.

3.3.2 | Influence of domain-general functions on basic number processing in PD (H2)

The question of whether numerical task performance can be explained by performance on domain-general tests will be addressed with further ANCOVAs. ANCOVAs will be carried out twice—one in a task-based and once in an effect-based approach. These lines of analysis will be preceded by correlational analyses of numerical performance and numerical effect (differing significantly between groups for H1) with possible covariates (executive functions, working memory, attention, verbal memory, language or visuospatial function, and motor symptoms and depression). For each numerical task/numerical effect (differing between groups), the cognitive covariate with the highest association will be included in the respective model, adding up to a maximum of three covariates including confounders identified for H1. For the task-based approach, this preliminary step is followed by separate ANCOVAs with the factor group (HC, PD-NC, or level I-PD-MCI) and the respective covariates for each numerical task. These ANCOVAs will be conducted for ACC (transcoding and (nonsymbolic number magnitude comparison) or percent absolute error (number line estimation). For the effect-based approach, this preliminary step is followed by separate ANCOVAs with the factor group (HC, PD-NC, or level I-PD-MCI) and the respective covariates for each numerical task. These ANCOVAs will be conducted on RTs for the respective effects. Following this frequentist approach, Bayesian ANCOVAs will be calculated to allow for further conclusions on the presence or absence of effects.

3.3.3 | Discrimination between cognitive statuses of PD patients by the tests of basic numerical cognition (H3)

The last diagnostic research question will be answered using logistic regression. This discriminant analysis with multiple predictors will be conducted on z-standardized performance scores for tasks differing between groups (i.e., transcoding, number line estimation, nonsymbolic, and/or symbolic number magnitude comparison) as well as significant numerical effects (i.e., Weber fraction, distance effect, and/or unit-decade compatibility effect) with the dependent variable of cognitive status (HC, PD-NC, or level I-PD-MCI). Covariates included in the preceding analyses will also be included in the model.
(i.e., a maximum of six clinical confounders and four cognitive covariates). The probability for members of the three groups to be (correctly or falsely) assigned to each group will be calculated based on the regression as a measure of prognostic value.

3.4 Possible limitations

The study design of the current project cannot control for all confounding effects coming from heterogeneous clinical samples. Differences in disease duration, medication, and staging of PD severity (Hoehn and Yahr staging) between the groups could limit the scope of the study, which is why variables differing between study groups will be included as covariates in the statistical model. However, interdependences between clinical covariates such as LEDD, motor symptoms, and disease duration need to be considered, and we will limit the number of clinical covariates to six in order to obtain a reasonable complexity of statistical models. As all PD patients will be tested in medication “on” state after intake of the regular dopaminergic medication, no major influence on patient’s performance is anticipated. Although we do not anticipate effects, we also think that including patients with medication has higher ecological validity, because the clinical reality is that PD patients, who are well diagnosed and investigated, get dopaminergic medication if this is deemed helpful. PD patients are also prone to cognitive fluctuations in association with attention. To limit the effect of attention deficits, critical tests will be administered at the beginning of sessions with breaks in-between to address patient attrition.

The fact that PD natural disease course is heterogeneous suggests analyzing clinical subtypes. First of all, the issue of patient groups differing too much will be addressed by potentially controlling for the UPDRS-III score as a covariate in the current study. The sample we want to test consists of a patient group as it presents in the clinic, inducing limited generalizability as patients might represent different PD subtypes. As the proportion of the respective subtypes will potentially be very small and we lack specific hypotheses for numerical cognition, a confirmatory analysis in line with open science and replicability principles would be difficult to achieve. However, following the Editor’s advice, we will “look into subtypes or biotypes (...) (even if preliminary and potentially underpowered, which could also be a specific limitation)” in an exploratory subtype analysis, investigating the effects of disease stage based on Hoehn and Yahr staging, motor type (Stebbins et al., 2013), and/ or age of onset defining common PD subgroups. These findings on heterogeneous subgroups can provide information to guide future research.

Furthermore, we try to limit PD heterogeneity by using the diagnosis of idiopathic PD and an age of 60 years or more as inclusion criteria. Other PD subtypes probably included in our study sample will be analyzed in exploratory analyses (e.g., stratifying by age of disease onset, Hoehn and Yahr staging, and/ or motor type), because our sample size affects statistical power for subgroup analyses and consequently hinders confirmatory analyses of small subsamples. We will also control for covariates differing between the two PD groups by including them in the ANCOVAs.

Furthermore, generalizability of our study is limited due to cultural specificities of basic number cognition—for example, studies with children show that performance in German samples is affected by the inversion property of the German multi-digit number system when compared to other (European or Western) cultures (Göbel et al., 2014). Generalizability to as well as comparability with other neurodegenerative diseases are also limited. However, it is not feasible to include other patient control groups such as AD patients in the current study design. The comparison with other neurodegenerative diseases is an interesting next step to generalize or differentiate number deficits across various neurodegenerative diseases, after we identified a systematic pattern of number cognition deficits in PD. As evidence on PD is rare, but research on numerical deficits in AD already has a tradition, the crucial and recent task is to systematically investigate PD before trying to compare it with other neurological diseases.

In the current study, we do not correct for multiple comparisons, because the Bayesian approach allows us to evaluate the effect sizes independent of frequentist corrections of the α-level. Future studies can then use our study’s effect sizes for power analyses and sample size estimation.

Psychometric flaws in clinical questionnaires are an alarming issue. It has already been shown that the UPDRS-III, the BDI-II (not in a PD-specific group), the FAQ, and the PDQ-39 do show psychometric issues when trying to apply Rasch modeling (Franchignoni et al., 2008; Hagell, 2019; Hagell & Nygren, 2007; Kirsch-Darrow et al., 2011; Siegert et al., 2010; Tosin et al., 2020; Wang & Gorenstein, 2013; Wu & Huang, 2010). When identifying these issues, the methodologically sound procedure would be to exclude problematic items. While this approach is psychometrically motivated, it is highly problematic for the clinical-neuropsychological application for several reasons: All these established measures assess different facets of the respective construct, with single items standing for different clinical symptoms. When excluding items, we would homogenize the heterogeneous symptoms, which are the very part of the nature of Parkinson’s disease. Neuropsychological instruments are validated for these heterogeneous constructs they are intending to measure and established in clinical research practice. Accordingly, the clinical scales can only be interpreted meaningfully when assessing them in their entirety. The aim of our study is to identify the clinical profile of a PD patient at risk of numerical impairment. As this profile needs to be generalizable to the clinical use, we need to score standard clinical measures as established in the clinical routine. However, we could evaluate the psychometric properties of the critical instruments in an exploratory analysis and to report them in the Supporting Information (https://osf.io/ap6je/).

Another psychometric issue is the assessment of language skill. The guideline by the Movement Disorder Society on the PD-MCI diagnosis (Litvan et al., 2012) suggests the Boston Naming test, the WAIS-IV subtest similarities (Petermann, 2012), and the Graded
Naming Test (McKenna & Warrington, 1983). However, all of them show methodological constraints: The Boston Naming test displays a ceiling effect and poor psychometric properties (Harry & Crowe, 2014), the WAIS-IV subtest similarities are confounded with executive function, and the Graded Naming Test does not exist in a German adaptation and is confounded with linguistic and cultural attributes. Therefore, we opted for the WAIS-IV subtest similarities, because we consider it most suited for our needs regarding difficulty and cultural standardization. However, we want to critically acclaim that it does not only assess language skills, but also assess the facets of executive function. Consequently, future work on a methodologically sound neuropsychological test of language skill in German PD patients based on patholinguistic expertise would be greatly desired for our scientific practice.

If our first study verifies deficits in number cognition in PD, further studies in larger samples might build on our behavioral findings by collecting biomarkers (blood, CSF, imaging) to identify the underlying pathomechanisms. In the current study, we are interested in numerical deficits occurring under regular intake of dopaminergic medication, to make inferences for an ecologically valid sample and with relevance for daily living. However, it would be interesting for follow-up studies to investigate the pathomechanisms by addressing the influence of PD medication (i.e., comparison between on- and off-state) on numerical deficits.

Finally, our target group most likely has been diagnosed before the most recent diagnostic criteria from the movement disorder society (Postuma & Berg, 2017; Postuma et al., 2016, 2018) were put into practice. Therefore, diagnosis will be based on the U.K. Parkinson’s Disease Society Brain Bank clinical diagnostic criteria (Hughes et al., 1992), despite their diagnostic inferiority because the MDS criteria cannot be applied retrospectively. Due to the complex and time-consuming design of the study, we had to opt for the diagnosis “level I-PD-MCI based on cognitive screening” using the MoCA to define patients’ global cognitive status. However, comprehensive test batteries allow for a more detailed evaluation of cognitive status, also differentiating single- and multi-domain PD-MCI. When assessing the progression of numerical deficits in association with cognitive decline, future studies should use a comprehensive test battery to explore the association between cognitive status and numerical deficits. Based on basic numerical deficits identified in the current study, future studies should assess the influence of numerical deficits on patients’ everyday life.

3.5 | Potential unexpected outcomes

Heterogeneous patient profiles might lead to ambivalent results when averaging over participants in our analyses. To address the heterogeneity, we might explore the data with subject-by-subject analyses with individual regression slopes to identify effect sizes per participant. These individual error profiles might show the patterns of co-occurring or distinct errors. Additionally, the taxonomy for transcoding errors was adapted from children and Alzheimer’s studies, so it might be necessary to extend it in case of Parkinson-specific transcoding errors not categorized in the literature so far. Furthermore, numerical deficits might be associated with specific subgroups of PD patients (e.g., gender effects; Loenneker et al., 2020) which might be addressed in exploratory analyses.

4 | FURTHER PROCEDURES

Testing is planned for after critical revisions of the preregistered review and following in principle acceptance. However, experiments can only start when the current COVID-19 situation permits human-to-human testing with elderly participants. Recruitment and testing phases are estimated to take 9 months. Data analysis and preparation of the final manuscript are expected to be finished 4 months after the last experimental session.

ACKNOWLEDGMENTS

HL received scholarships of the Landesgraduiertenfoerderung Baden-Wuerttemberg and of the Friedrich-Ebert Stiftung. HL was granted funding by the Wilhelm Schuler Foundation covering participants’ reimbursement and testing material with a grant over 3853.50€. CA was supported by the European Social Fund and by 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 thank Julia Huber, Krzysztof Cipora, and Mojtaba Soltanlou for methodological and statistical advice, and Zoe Kirste for language proofreading.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All the authors had 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, H.D.L., C.A., H.C.N., and I.L.S.; Data Curation, H.L.; Methodology, H.L., C.A., H.C.N., K.W., and I.L.S.; Investigation, H.L.; Formal Analysis, H.L.; Writing – Original Draft, H.L.; Writing – Review & Editing, H.L., C.A., H.C.N., K.W., and I.L.S.; Project Administration, H.L.; Visualization, H.L.; Funding Acquisition, H.L.
REFERENCES

Aarsland, D., Creste, B., Politis, M., Chaudhuri, K. R., Ffythe, 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

Aarsland, D., Larsen, J. P., Tandberg, E., & Laake, K. (2000). Predictors of nursing home placement in Parkinson’s disease: A population-based, prospective study. Journal of the American Geriatrics Society, 48(8), 938–942. https://doi.org/10.1111/j.1532-5415.2000.tb06891.x

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

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

Barker, R. A., & Williams-Gray, C. H. (2014). Mild cognitive impairment and Parkinson’s disease—Something to remember. Journal of Parkinson’s Disease, 4(4), 651–656. https://doi.org/10.2233/jpd-JPD-140427

Becker, S., Boettinger, O., Sulzer, P., Hobert, M. A., Brockmann, K., Maetzler, W., Berg, D., Liepelt-Scarfone, I., & Gifford, K. (2021). Everyday function in Alzheimer’s and Parkinson’s patients with mild cognitive impairment. Journal of Alzheimer’s Disease, 79(1), 197–209. https://doi.org/10.3233/JAD-200256

Benton, A. L., Varney, N. R., & Hamsher, K. d. S. (1978). Visuospatial judgement in Alzheimer’s and Parkinson’s disease 53 years on. The British Journal of Psychiatry, 133(5), 276–282. https://doi.org/10.1192/bjp.133.5.276

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., Del Tredici, K., Rüb, U., de Vos, R. A., Jansen Steur, E. N., & Baayen, H. R., & Milin, P. (2010). Analyzing reaction times. Psychological Bulletin, 136(1), 155–159. https://doi.org/10.1037/a0018816

Cipora, K., Schroeder, P. A., Soltanlou, M., & Nuerk, H. (2018). More space, better math: Is space a powerful tool or a cornerstone for understanding arithmetic? In K. S. Mix & M. T. Battista (Eds.), Visualizing mathematics: The role of spatial reasoning in mathematical thought (pp. 77–116). Springer. https://doi.org/10.1007/978-3-319-98767-5_4

Dietrich, J. F., Huber, S., & Nuerk, H.-C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS)–A research review. Frontiers in Psychology, 6, 295. https://doi.org/10.3389/fpsyg.2015.00295

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, C. 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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. Journal of Cognitive Neuroscience, 15(1), 47–56. https://doi.org/10.1162/089892903321107819

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
Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer’s & Dementia*, 9(1), 63–75.e2. https://doi.org/10.1016/j.jalz.2012.11.007

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rinne, J. O., Portin, R., Ruottila, 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

Roheger, M., Kalbe, E., & Liepelt-Scarfone, I. (2018). Progression of cognitive decline in Parkinson’s disease. *Journal of Parkinson’s Disease*, 8(2), 183-193. https://doi.org/10.3233/JPD-181306

Rosselli, M., Ardila, A., Arvizu, L., Kretzner, T., Standish, V., & Liebermann, J. (1998). Arithmetical abilities in Alzheimer disease. *International Journal of Neuroscience*, 96(3-4), 141-148. https://doi.org/10.3109/00207459808986463

Savica, R., Grossardt, B. R., Rocca, W. A., & Bower, J. H. (2018). Parkinson’s disease with and without Dementia: A prevalence study and future projections. *Movement Disorders*, 33(4), 537-543. https://doi.org/10.1002/mds.27277

Schretlen, D., Pearlson, G. D., Anthony, J. C., Aylward, E. H., Augustine, A. M., Davis, A., & Barta, P. (2000). Elucidating the contributions of processing speed, executive ability, and frontal lobe volume to normal age-related differences in fluid intelligence. *Journal of the International Neuropsychological Society*, 6(1), 52–61. https://doi.org/10.1017/S1355617700611062

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). Neuropsychometric predictors of financial capacity across the dementia spectrum: Normal aging, mild cognitive impairment, and Alzheimer’s disease. *Journal of the International Neuropsychological Society*, 15(2), 258–267. https://doi.org/10.1017/S1355617709090365

Siegent, R. J., Tennant, A., & Turner-Stokes, L. (2010). Rasch analysis of the beck depression inventory-II in a neurological rehabilitation sample. *Disability and Rehabilitation*, 32(1), 8–17. https://doi.org/10.3109/09638280902971398

Simpson, J. R. (2014). DSM-5 and neurocognitive disorders. *Journal of the American Academy of Psychiatry and the Law*, 42(2), 159–164.

Soltanlou, M., Pixner, S., & Nuerk, H.-C. (2015). Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: A longitudinal investigation. *Frontiers in Psychology*, 6, 1062. https://doi.org/10.3389/fpsyg.2015.01062

Stebbins, G. T., Goetz, C. G., Burn, D. J., Jankovic, J., Khoo, T. K., & Tilley, B. C. (2013). How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson’s disease rating scale: Comparison with the unified Parkinson’s disease rating scale. *Movement Disorders*, 28(5), 668–670. https://doi.org/10.1002/mds.25383

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. https://doi.org/10.3233/JAD-180080

Tolminson, 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

Tosin, M. H. D. S., Goetz, C. G., Luo, S., Choi, D., & Stebbins, G. T. (2020). Item response theory analysis of the MDS-UPDRS motor examination: Tremor vs. nontremor Items. *Movement Disorders*, 35(9), 1587–1595. https://doi.org/10.1002/mds.28110

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

Uribe, C., Segura, B., Baggio, H. C., Camapabadal, A., Abos, A., Compta, Y., Martí, M. J., Valdeolrios, 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

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria*, 35(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience & Biobehavioral Reviews*, 57, 209–219. https://doi.org/10.1016/j.neubiorev.2015.09.005

Wu, P.-C., & Huang, T.-W. (2010). Person heterogeneity of the BDI-II-C and its effects on dimensionality and construct validity: Using mixture item response models. *Measurement and Evaluation in Counseling and Development*, 43(3), 155–167. https://doi.org/10.1177/0748175610384808

Zamarian, L., Visani, P., Delazer, M., Seppi, K., Mair, K. J., Diem, A., Poewe, W., & Benke, T. (2006). Parkinson’s disease and arithmetic: 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

Zimmermann, P., & Fimm, V. (2017). TestbatteriezurAufmerksamkeitsprüfung (TAP) (2.3.1). Psychologische Testsysteme.

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology*, 102(1), 60–77. https://doi.org/10.1016/j.jecp.2008.04.003