

Short title: Cognitive summary score in Parkinson's disease

## **Use of robust norming to create a sensitive cognitive summary score in *de novo* Parkinson's disease**

Daniel Weintraub, MD<sup>1,2</sup> Michael C. Brumm, MS<sup>3</sup>, Ryan Kurth, MS<sup>3</sup>, Michele K. York, PhD, ABPP-CN<sup>4</sup> and the Parkinson's Progression Markers Initiative\*

<sup>1</sup>Departments of Psychiatry and Neurology, Perelman School of Medicine at the University of Pennsylvania

<sup>2</sup>Parkinson's Research, Education and Clinical Center, Philadelphia Veterans Affairs Medical Center

<sup>3</sup>Department of Biostatistics, College of Public Health, University of Iowa

<sup>4</sup>Departments of Neurology and Psychiatry and Behavioral Sciences, Baylor College of Medicine

\*See Supplementary Table\_PPMI Authorship List

Abstract: 346 words

Manuscript: 3007 words

References: 29

Tables: 4 + 1 supplementary

Figures: 1

Funding: Michael J. Fox Foundation for Parkinson's Research (Research Grant "Creation and Performance of a Composite Cognitive Test Score in Parkinson Disease"; Grant ID: MJFF-022598)

Acknowledgments: PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, Abbvie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, BioArctic, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Capsida Biotherapeutics, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Jazz Pharmaceuticals, Johnson & Johnson Innovative Medicine, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Neuron23, Neuropore, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation and Yumanity Therapeutics.

### Author ORCID IDs

Weintraub: 0000-0003-0633-7168

Brumm: 0000-0002-3173-9989

Kurth: 0009-0004-4569-5763

York: 0000-0003-4056-0742

Short title: Cognitive summary score in Parkinson's disease

## ABSTRACT

**Background and Objectives:** Cognitive impairment is common at all stages in Parkinson's disease (PD). However, the field is hampered by consensus over which neuropsychological tests to use and how to utilize the results generated by a cognitive battery. An option that combines the richness of a neuropsychological battery with the simplicity of a single test score is a cognitive summary score (CSS). The objective was to determine if a CSS created using robust norming is sensitive in detecting early cognitive deficits in *de novo*, untreated PD.

**Methods:** Using baseline cognitive data from PD participants and healthy controls (HCs) in the Parkinson's Progression Markers Initiative, these steps were taken: (1) creating a robust HC subgroup that did not demonstrate cognitive decline over time; (2) using the robust HC subgroup to create regression-based internally-derived standardized scores (z-scores) for six cognitive scores across five tests; and (3) creating a CSS by averaging all standardized test z-scores.

**Results:** PD participants scored worse than HCs on all cognitive tests, with a larger effect size (PD versus HCs) when the comparison group was the robust HC subgroup compared with all HCs. Applying internally-derived norms rather than published norms, the largest cognitive domain effect sizes (PD vs. robust HCs) were for processing speed/working memory (Cohen's  $d = -0.55$ ) and verbal episodic memory (Cohen's  $d = -0.48$  and  $-0.52$ ). In addition, using robust norming shifted PD performance from the middle of the average range (CSS z-score =  $-0.01$ ) closer to low average (CSS z-score =  $-0.40$ ), with the CSS having a larger effect size (PD vs. robust HC subgroup; Cohen's  $d = -0.60$ ) compared with all individual cognitive tests.

**Discussion:** PD patients perform worse cognitively than HC at disease diagnosis on multiple cognitive domains, particularly information processing speed and verbal memory. Using robust norming increases effect sizes and lowers the scores of PD patients to "expected" levels. The CSS performed better than all individual cognitive tests. A CSS developed using a robust norming process may be sensitive to cognitive changes in the earliest stages of PD and have utility as an outcome measure in clinical research, including clinical trials.

Short title: Cognitive summary score in Parkinson's disease

## INTRODUCTION

Long-term significant cognitive impairment is a common and dreaded outcome in Parkinson's disease (PD), with dementia affecting up to 80% of patients long-term<sup>1,2</sup>. Mild cognitive impairment (MCI) is also relatively common<sup>3</sup>, even in *de novo* or early disease<sup>4-6</sup>. In addition, subtle cognitive changes can be detected in the prodromal state, including in patients with isolated REM sleep behavior disorder (iRBD)<sup>7</sup> or persons with hyposmia<sup>8</sup>.

However, the ability to track cognitive changes over time, including in the context of short-term randomized controlled trials (RCTs), is hampered by consensus over which neuropsychological tests to use and how to utilize the results<sup>9,10</sup>. This may have contributed to having only one cognitive enhancing medication FDA-approved for the treatment of PD dementia (rivastigmine)<sup>11</sup> with no other positive RCTs in the 20 years since then, and no positive RCTs for PD-MCI<sup>12</sup>.

Increasingly computerized cognitive batteries and smartphone-based cognitive apps are being used in PD, but most are not yet fully validated and ready for routine use in clinical research. Global screening instruments commonly used in PD (e.g., Mini-Mental State Examination [MMSE]<sup>13</sup> and Montreal Cognitive Assessment [MoCA])<sup>14</sup> are brief (i.e., 5-10 minutes), but may not be sensitive to mild cognitive changes in the short-term or may have significant variability.

Many individual paper-and-pencil cognitive tests are used in PD, both in clinical care and clinical research, and they are commonly assembled into a cognitive battery. Given the heterogeneity of cognitive deficits in PD, even at the stage of MCI<sup>3</sup>, batteries often assess multiple cognitive domains. Many versions of detailed cognitive batteries (e.g., recommended Level I or Level II PD-MCI batteries<sup>9</sup>) are utilized, with some evidence that there is differential sensitivity among commonly-used neuropsychological tests to mild cognitive deficits in PD<sup>15</sup>. Recent research has suggested that in general Level I and Level II cognitive batteries are better able to predict conversion from PD-MCI to PD dementia than are global screening instruments<sup>16</sup>. However, it remains unclear how to best utilize and interpret results containing multiple individual (sub)scores from differently-scaled and -normed tests.

An option that allows one to combine the richness and details of a neuropsychological battery with the simplicity of a single test score is to create a cognitive summary score (CSS) after all detailed test results are normed on the same scale (e.g., z-score). Examples from other neurodegenerative diseases include: (1) Alzheimer's disease (AD): the Preclinical Alzheimer Cognitive Composite (PACC; consisting of total recall score from the Free and Cued Selective Reminding Test, delayed recall score on the Logical Memory IIa sub-test from the Wechsler Memory Scale, Digit Symbol Substitution Test and MMSE total score)<sup>17</sup>, and the AD Composite Score (ADCOMS; 4 Alzheimer's Disease Assessment Scale—cognitive subscale items, 2 Mini-Mental State Examination items, and all 6 Clinical Dementia Rating—Sum of Boxes items)<sup>18</sup>; and (2) Huntington's disease (HD): the Huntington's Disease Cognitive Assessment Battery (HD-CAB; consisting of Symbol Digit Modalities Test, Paced Tapping, One Touch Stockings of Cambridge (abbreviated), Emotion Recognition, Trail Making B and Hopkins Verbal Learning Test)<sup>19</sup>.

In the Parkinson Associated Risk Syndrome (PARS) study a CSS was generated from a very detailed cognitive battery and used to document subtle cognitive changes in prodromal PD (i.e., persons with hyposmia + dopamine transporter [DAT] SPECT scan deficit)<sup>8,20</sup>. To our knowledge, there have not been other published literature documenting attempts to create a cognitive composite from detailed individual tests.

## Short title: Cognitive summary score in Parkinson's disease

If determined to be more sensitive to early cognitive decline in PD, it is possible that a CSS would allow more neurobiological predictors of cognitive decline to be identified, and even serve as an outcome measure for RCTs, including in persons with prodromal or at-risk PD. In addition, such a CSS could be valuable in the routine, clinical evaluation of PD patients, as clinical neuropsychological evaluations often administer the battery of standard neuropsychological tests utilized in PPMI.

The Parkinson's Progression Markers Initiative (PPMI) study originally had a PD-MCI Level I battery (five detailed, well-established cognitive tests) with published norms<sup>21</sup>. Given that PPMI PD participants are highly educated and motivated, we implemented robust norming to increase sensitivity to detect cognitive differences in PD compared with healthy controls (HCs), in a fashion similar to that employed to update the norms for the Dementia Rating Scale<sup>22</sup>. We hypothesized that using robust norming would increase sensitivity and adjust PD participant scores to the expected slight impairment reported in the literature<sup>6</sup>, and that a cognitive summary score would be more sensitive than individual test results in detecting cognitive differences in *de novo*, untreated PD patients compared with HCs.

## **METHODS**

### **Standard protocol approvals, registrations and patient consents**

An ethical standards committee on human experimentation reviewed and approved the study at each site. Written informed consent was obtained from all participants in the study.

### **Cohort**

The PPMI study and cohort has been extensively described<sup>23,24</sup>. All participants signed an approved informed consent form. Inclusion criteria for PD participants included: (1) an asymmetric resting tremor or asymmetric bradykinesia, or at least two symptoms out of resting tremor, bradykinesia, and rigidity; (2) a recent clinical diagnosis of PD (mean [SD] duration from diagnosis = 8.5 [7.3] months); (3) being untreated with PD medications; (4) evidence of dopaminergic deficit based on DAT SPECT imaging; and (5) being non-demented. HCs were required to (1) not have clinically significant neurological dysfunction, (2) not have a first-degree relative with PD and (3) have a Montreal Cognitive Assessment (MoCA) score  $\geq 27$ . Participants enrolled as HCs who had evidence of CSF  $\alpha$ -synuclein seed amplification assay (SAA) positivity were not considered to be HCs for the purposes of this analysis.

### **Creating database**

Baseline data were used, at which point PD participants were recently diagnosed and had not yet begun treatment with PD medication. The original cognitive battery was used to create the CSS to optimize the amount of data available for analyses. This original battery was composed of the Hopkins Verbal Learning Test-Revised (HLVT-R immediate and delayed free recall scores)(24), the Benton Judgment of Line Orientation - 15 item version (JLO)(25), Symbol-Digit Modalities Test (SDMT)(26), Letter-Number Sequencing (LNS)(27), and semantic (animal) fluency(28). Together these tests assess memory, visuospatial function, information processing speed, executive function, working memory and language.

### **Published norms**

Short title: Cognitive summary score in Parkinson's disease

Using previously published methods derived from participants not enrolled in PPMI, “external” standard scores were derived. This included T-scores for the HVLt-R (immediate and delayed free recall scores), SDMT and semantic fluency, and scaled scores for the JLO and LNS.

### **Robust norming**

The robust HC group was defined using the criteria depicted in **Figure 1**. Briefly, participants were required to have a baseline MoCA of  $\geq 27$ , have a year 1 MoCA score of  $\geq 26$ , and not have had more than a 2-point drop in their MoCA score between baseline and year 1. This subset of participants was used to create internal norms for each test which were then used to calculate internally-derived z-scores for each participant (see Statistical Analysis for details).

### **Data Availability**

PPMI data used in the preparation of this article were obtained November 6, 2023 from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/access-data-specimens/download-data](http://www.ppmi-info.org/access-data-specimens/download-data)), RRID:SCR\_006431. This analysis was conducted by the PPMI Statistics Core and used actual dates of activity for participants, a restricted data element not available to public users of PPMI data. PPMI data are publicly available from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/access-data-specimens/download-data](http://www.ppmi-info.org/access-data-specimens/download-data)). For up-to-date information on the study, visit [www.ppmi-info.org](http://www.ppmi-info.org).

### **Statistical Analysis**

Statistical analyses were performed using SAS v9.4 (SAS Institute Inc., Cary, NC; [sas.com](http://sas.com); RRID:SCR\_008567). Statistical analysis codes used to perform the analyses in this article are shared on Zenodo [[10.5281/zenodo.12636045](https://doi.org/10.5281/zenodo.12636045)].

All external standard scores derived from published norms (T-scores or scaled scores) were converted to z-scores and then capped at -3 to 3. To derive internal norms, the robust HC population was used to build a linear regression model for each test. Age, sex, and continuous education were considered as possible predictors of raw test scores. The second order polynomials of centered age and education were also considered. Sex was coded as a binomial variable with male = 1 and female = 2. Backwards selection with a cutoff of  $\alpha = 0.10$  was used to determine which variables would appear in each final model. The Kolmogorov–Smirnov test was used to assess the normality of the residuals of each model. These derived linear models were used to calculate an “expected” raw score for each participant and each test, and then a z-score was calculated by taking the participant's raw score minus their expected score and dividing by the root mean square error. These internal norms were also capped at -3 to 3.

Two cognitive summary scores were calculated for each participant, one being an average of their six external z-scores (derived from published norms), and the other an average of their six internal z-scores (derived from the PPMI robust HC subgroup), with the six z-scores weighted equally in both cases.

For each variable in Table 1, two-sample t-tests and chi-squared tests were performed to compare the PD group to each of the HC groups. Cohen's d effect sizes were used to compare these composite scores across groups.

Short title: Cognitive summary score in Parkinson's disease

## RESULTS

### Cohort characteristics

There were 923 participants with *de novo*, untreated PD; they had mean disease duration <1 year, mean age of 63 years, and were two-thirds male, highly educated, and overwhelmingly white and non-Hispanic (**Table 1**). There were 250 total HCs, and 154 participants in the robust HC subgroup. PD participants were older than the total HCs (mean age = 61) and robust HCs (mean age = 59), but otherwise had similar demographic characteristics.

### Raw cognitive test scores

PD participants scored worse on global cognition (MoCA score) by approximately 1 point compared with both HC groups, as well as worse on all detailed cognitive test raw scores, except LNS for the entire HC group (**Table 1**). The entire HC group scored worse numerically on all cognitive tests compared with the robust HC subgroup.

### Using robust HC subgroup instead of entire HC group

PD participants scored worse than the entire HC group and the robust HC subgroup on all detailed cognitive tests using both published norms (**Table 2**) and internally-derived norms (**Table 3**). The effect size (Cohen's *d*) was greater (sometimes twice as great) when comparing the PD group to the robust HC subgroup rather than the entire HC group for all cognitive tests, whether using published norms (**Table 2**) or internally-derived norms (**Table 3**). The tests with the biggest effect sizes, which were medium-sized, were for information processing speed (SDMT) and verbal episodic learning and memory (HVLTR). Smaller effect sizes were seen for JLO, LNS and semantic fluency, but the effect size was increased for these tests by using the robust HC subgroup as the comparator group.

### Regression-based norms derived from robust healthy controls

The results of regression analyses of the effects of age, sex, and education on the raw test scores of robust SH participants are summarized in **Supplementary Table 1**. In brief, significant predictors included age and sex for the immediate and delayed recall scores of the HVLTR; age, sex, and education for the JLO; age and education for the LNS; only sex only for semantic fluency; and only age for the SDMT. For all tests where age was a significant predictor, older participants scored lower on average. Females scored significantly higher on the HVLTR and semantic fluency; whereas males scored higher on the JLO. Participants with higher education performed better on the JLO and LNS. Kolmogorov–Smirnov tests (results not shown) indicated normally distributed residuals for HVLTR immediate recall, LNS, SDMT, and semantic fluency; however, residuals were left-skewed for HVLTR delayed recall (skewness = -0.80; *D* = 0.09; *p* < 0.01) and JLO (skewness = -1.36; *D* = 0.11; *p* < 0.01).

### Using internally-derived norms instead of published norms

The effect sizes for PD versus both all HC and the robust HC subgroup were similar for internally-derived (**Table 3**) vs. published norms (**Table 2**), except for semantic fluency, for which the effect size was approximately twice as great when using internally-generated norms. The main change occurring with internally-derived norms was a shift for nearly all tests (except HVLTR delayed recall, which was relatively unchanged) to a more negative (i.e., more impaired) z-score, for both PD participants and HCs. Specifically, z-scores for PD participants

Short title: Cognitive summary score in Parkinson's disease

ranged from -0.50 to +0.65 using published norms, but -0.56 to -0.22 using internally-derived norms.

### **Cognitive summary score**

Using published norms, the CSS showed a larger effect size than all individual tests for both HC groups, except for the SDMT (**Table 4**). Using internally-derived norms, the CSS had a larger effect size than all individual tests when using the robust HC subgroup as the comparator group (**Table 4**). Once again, changing from published to internally-derived norms decreased the z-score for the CSS from average (-0.01) to closer to low average (-0.40).

### **DISCUSSION**

Using PPMI baseline cognitive data from *de novo*, untreated PD patients and demographically-similar HCs, we demonstrated that it is possible to create a CSS using robust norming that is sensitive to subtle cognitive changes in early PD.

Robust norming has been used to update the norms of commonly-used cognitive tests<sup>25</sup>, given that studies have indicated that cross-sectional normative samples of older adults include a mix of cognitively normal individual but also those in the initial stages of cognitive decline. This contaminates cross-sectional normative samples with cases of preclinical dementia that leads to an underestimation of the test mean and overestimation of the variance, thus reducing their clinical utility. When the DRS was renormed, use of the robust norms resulted in an almost twofold increase in sensitivity with relatively minimal loss of specificity when compared with the original norms<sup>22</sup>.

Cognitive summary scores, also call cognitive composites, have been developed and utilized in clinical research for both AD and HD, including RCTs. The ADCOMS, developed to be sensitive to the early stages of AD and to changes with symptomatic therapy<sup>18</sup>, showed statistically significant improvement with lecanemab treatment in a recent AD trial<sup>26</sup>. The PACC has been shown to be sensitive to beta amyloid (A $\beta$ ) positivity in cognitively-unimpaired individuals<sup>27</sup>, although it was negative as the primary outcome in a trial of solanezumab for preclinical AD<sup>28</sup>. For HD, the HD-CAB, which yields a composite z-score that is sensitive in pre-HD and early HD group, and shows high test-retest reliability<sup>19</sup>, is recommended for use in clinical care and research, but may need modifications and there aren't published results reporting on its use in a RCT<sup>29</sup>.

We found that PD patients performed worse than HCs, whether using raw scores, published norms or internally-derived norms. This is consistent with what has been reported in other studies of newly-diagnosed PD cohorts<sup>5</sup>, and with research showing that 15-30% of *de novo* patients meet criteria for MCI<sup>6</sup>.

PD patients also performed worse cognitively than HCs on all cognitive domains assessed, which is also consistent with multiple domains being impaired at the stage of MCI in PD<sup>3</sup>, which makes the case that any cognitive battery in PD should assess multiple cognitive domains. Regarding specific cognitive tests, the most notable differences when comparing PD patients with HCs was for SDMT (information processing speed) and HVLT-R (verbal learning and memory), suggesting that these tests are good candidates for inclusion in a PD cognitive battery.

Short title: Cognitive summary score in Parkinson's disease

As part of robust norming, we excluded HCs who experienced cognitive decline over the first year of the study or had evidence for neuronal alpha-synuclein disease based on their CSF alpha-synuclein seed amplification assay test. Using the robust HC subgroup instead of the entire HC group significantly increased the effect sizes of cognitive differences between PD patients and HCs, confirming the value in excluding HCs with incipient cognitive decline from the norming process.

The second stage of the robust norming process was the creation of internally-derived norms instead of relying on published norms obtained from different populations, which is important given that the PPMI cohort, both PD patients and HCs, are highly educated (82% of the PD patients reported having formal education beyond high school, and only two cognitive tests in the battery adjust for education), motivated and younger than some other newly-diagnosed cohorts. Using the internally-derived versus published norms lowered the standardized test scores of PD patients to “expected” levels (between average and low average), and allowed direct comparison across the tests due to the same reference population being used, as opposed to different reference populations for most published norms.

The final step after robust norming was to create a CSS by averaging the individual test z-scores generated by robust norming. When doing this, and using the robust HC subgroup as the comparator, the cognitive summary score had a greater effect size than any of the individual tests comprising the battery, showing enhanced sensitivity for a summary score when using a robust HC group. This process also allows for applying z-score cut-offs to define cognitive impairment, which might be used as an inclusion criterion or as an endpoint in clinical trials.

Future research for a cognitive summary score in PD might include using other cognitive tests to compare performance against this battery, testing it against a global screening instrument (e.g., the MoCA), considering unequal weighting of individual test scores based on their effect size, testing it in prodromal PD (e.g., iRBD or hyposmia), determining how the cognitive summary score evolves and predicts outcomes over time, determining its association with biological variables and using it as an outcome measure in clinical trials.

Strengths of the study include a very large sample size, a multi-site international cohort, and use of well-validated and commonly-used neuropsychological tests. Limitations include not having a Level II PD-MCI cognitive battery and having different inclusion criteria at baseline for PD patients and HCs (i.e., HCs had to have MoCA score  $\geq 27$ , while PD patients did not).

In conclusion, robust norming enhanced sensitivity in detecting cognitive differences in *de novo*, untreated PD, and generated scores consistent with the mild cognitive deficits known to commonly occur in prodromal and early disease. The CSS generated after the robust norming process combines the richness of a detailed, multi-domain cognitive battery with the simplicity a single, identically-normed and equally-weighted score with potential for domain subscores.



Short title: Cognitive summary score in Parkinson's disease

**Table 1. Demographic characteristics of PD and HC cohorts**

Variable	PD group	HC group		p-values	
	All <i>de novo</i> PD (N = 923)	All HC (N = 250)	Robust HC (N = 154)	PD vs. All HC	PD vs. Robust HC
<b>Age at baseline</b> , mean (SD)	63.1 (9.4)	61.2 (11.6)	59.1 (11.6)	0.0077	<.0001
Median (min, max)	63.6 (33.5, 84.9)	62.8 (30.4, 85.3)	60.5 (30.6, 85.3)		
<b>Sex</b> , n (%)					
Male	607 (65.8%)	154 (61.6%)	94 (61.0%)	0.2212	0.2548
Female	316 (34.2%)	96 (38.4%)	60 (39.0%)		
<b>Education</b> , mean (SD)	16.0 (3.1)	16.1 (3.1)	16.1 (3.0)	0.7029	0.8769
Median (min, max)	16 (5, 32)	16 (6, 24)	16 (8, 24)		
<b>Race</b> , n (%)					
White	856 (93.3%)	221 (88.8%)	140 (91.5%)	0.0155	0.4051
Black or African American	18 (2.0%)	16 (6.4%)	8 (5.2%)		
Asian	14 (1.5%)	5 (2.0%)	2 (1.3%)		
Other	29 (3.2%)	7 (2.8%)	3 (2.0%)		
<b>Ethnicity (Hispanic or Latino)</b> , n (%)	32 (3.5%)	4 (1.6%)	2 (1.3%)	0.1288	0.1523
<b>Months since PD diagnosis</b> , mean (SD)	8.5 (7.3)	N/A	N/A	—	—
Median (min, max)	5.9 (0.4, 42.8)	N/A	N/A		
<b>MDS-UPDRS Part III</b> , mean (SD)	22.4 (9.7)	1.3 (2.1)	1.1 (1.8)	<.0001	<.0001
Median (min, max)	21 (3, 62)	0 (0, 13)	0 (0, 10)		
<b>MoCA</b> , mean (SD)	26.9 (2.4)	28.0 (1.5)	28.3 (1.1)	<.0001	<.0001
Median (min, max)	27 (15, 30)	28 (20, 30)	28 (27, 30)		
<b>HVLT-R Immediate Recall</b> , mean (SD)	24.3 (5.1)	25.8 (4.7)	27.0 (4.4)	<.0001	<.0001
Median (min, max)	24 (9, 36)	26 (13, 36)	27.5 (14, 36)		
<b>HVLT-R Delayed Recall</b> , mean (SD)	8.3 (2.7)	9.2 (2.4)	9.7 (2.2)	<.0001	<.0001
Median (min, max)	9 (0, 12)	10 (2, 12)	10 (3, 12)		
<b>Judgment of Line Orientation</b> , mean (SD)	12.8 (2.2)	13.1 (2.1)	13.3 (1.9)	0.0243	0.0079
Median (min, max)	13 (1, 15)	14 (2, 15)	14 (4, 15)		
<b>Letter Number Sequencing</b> , mean (SD)	10.5 (2.7)	10.9 (2.7)	11.3 (2.8)	0.0607	0.0009
Median (min, max)	11 (2, 20)	11 (2, 20)	11 (2, 20)		
<b>Semantic Fluency (Animals)</b> , mean (SD)	21.2 (5.3)	21.9 (5.4)	22.6 (5.5)	0.0470	0.0018
Median (min, max)	21 (7, 41)	21.5 (10, 38)	22 (10, 38)		
<b>Symbol Digit Modalities Test</b> , mean (SD)	41.9 (10.3)	47.3 (10.3)	48.8 (10.5)	<.0001	<.0001
Median (min, max)	42 (7, 82)	47 (20, 83)	48 (20, 83)		

Abbreviations: HC = healthy controls; HVLT-R = Hopkins Verbal Learning Test-Revised; MDS-UPDRS = Movement Disorder Society Unified Parkinson's Disease Rating Scale; MoCA = Montreal Cognitive Assessment; PD = Parkinson's disease.

Report Generated on Data Submitted as of: 06Nov2023. Among *de novo* PD cohort, ≤10 participants missed any one assessment; among HCs, ≤2 participants in overall group and ≤1 participant in the robust subgroup missed any one assessment.

Short title: Cognitive summary score in Parkinson's disease

**Table 2. Published norms z-scores at baseline among de novo PD and HC cohorts**

Variable	PD Group	HC Group		Cohen's d*	
	All <i>de novo</i> PD (N = 923)	All HC (N = 250)	Robust HC (N = 154)	PD vs All HC	PD vs Robust HC
<b>HVLT-R Immediate Recall</b> , mean (SD)	-0.42 (1.09)	-0.12 (1.04)	0.08 (1.02)	-0.28	-0.47
Median (min, max)	-0.40 (-3.00, 2.40)	0.00 (-3.00, 2.40)	0.10 (-3.00, 2.40)		
<b>HVLT-R Delayed Recall</b> , mean (SD)	-0.50 (1.18)	-0.10 (1.09)	0.06 (1.01)	-0.34	-0.48
Median (min, max)	-0.30 (-3.00, 3.00)	0.20 (-3.00, 1.40)	0.35 (-3.00, 1.40)		
<b>Judgment of Line Orientation</b> , mean (SD)	0.65 (0.99)	0.83 (0.94)	0.89 (0.90)	-0.18	-0.25
Median (min, max)	0.80 (-2.81, 2.49)	0.95 (-2.21, 2.57)	1.03 (-2.21, 2.57)		
<b>Letter Number Sequencing</b> , mean (SD)	0.52 (0.94)	0.61 (0.93)	0.71 (0.96)	-0.10	-0.20
Median (min, max)	0.67 (-2.67, 3.00)	0.67 (-2.67, 3.00)	0.67 (-2.67, 3.00)		
<b>Semantic Fluency (Animals)</b> , mean (SD)	0.10 (1.00)	0.17 (1.10)	0.25 (1.11)	-0.07	-0.14
Median (min, max)	0.10 (-3.00, 3.00)	0.10 (-2.80, 3.00)	0.20 (-2.60, 3.00)		
<b>Symbol Digit Modalities Test</b> , mean (SD)	-0.40 (0.98)	0.08 (1.02)	0.19 (1.03)	-0.49	-0.60
Median (min, max)	-0.38 (-3.00, 3.00)	0.00 (-2.75, 2.90)	0.11 (-2.38, 2.90)		

Abbreviations: HC = healthy controls; HVLT-R = Hopkins Verbal Learning Test-Revised; PD = Parkinson's disease.

Report Generated on Data Submitted as of: 06Nov2023. Among *de novo* PD cohort, ≤13 participants were missing a z-score for any one assessment; among HCs, 1 participant (not in the robust HC subgroup) had a missing z-score for Judgment of Line Orientation.

\* Cohen's d was computed by dividing the mean difference between groups by the pooled standard deviation.

Short title: Cognitive summary score in Parkinson's disease

**Table 3. Internally-derived, regression-based z-scores at baseline among de novo PD and HC cohorts**

Variable	PD Group	HC Group		Cohen's d*	
	All <i>de novo</i> PD (N = 923)	All HC (N = 250)	Robust HC (N = 154)	PD vs All HC	PD vs Robust HC
<b>HVLT-R Immediate Recall</b> , mean (SD)	-0.56 (1.11)	-0.25 (1.04)	0.00 (0.99)	-0.29	-0.52
Median (min, max)	-0.52 (-3.00, 2.21)	-0.18 (-3.00, 2.37)	0.06 (-2.88, 2.37)		
<b>HVLT-R Delayed Recall</b> , mean (SD)	-0.54 (1.14)	-0.17 (1.07)	0.00 (0.99)	-0.32	-0.48
Median (min, max)	-0.33 (-3.00, 1.56)	0.03 (-3.00, 1.61)	0.18 (-3.00, 1.61)		
<b>Judgment of Line Orientation</b> , mean (SD)	-0.24 (1.07)	-0.04 (1.01)	0.01 (0.95)	-0.18	-0.23
Median (min, max)	-0.03 (-3.00, 1.80)	0.18 (-3.00, 2.02)	0.19 (-3.00, 2.02)		
<b>Letter Number Sequencing</b> , mean (SD)	-0.22 (0.95)	-0.12 (0.94)	0.00 (0.99)	-0.10	-0.22
Median (min, max)	-0.20 (-3.00, 3.00)	-0.15 (-2.85, 3.00)	-0.10 (-2.85, 3.00)		
<b>Semantic Fluency (Animals)</b> , mean (SD)	-0.25 (0.96)	-0.12 (0.98)	0.00 (1.00)	-0.13	-0.26
Median (min, max)	-0.36 (-3.00, 3.00)	-0.18 (-2.31, 2.73)	-0.12 (-2.31, 2.73)		
<b>Symbol Digit Modalities Test</b> , mean (SD)	-0.56 (1.01)	-0.07 (0.98)	0.00 (0.98)	-0.49	-0.55
Median (min, max)	-0.56 (-3.00, 3.00)	-0.05 (-2.46, 3.00)	-0.04 (-2.46, 3.00)		

Abbreviations: HC = healthy controls; HVLT-R = Hopkins Verbal Learning Test-Revised; PD = Parkinson's disease.

Report Generated on Data Submitted as of: 06Nov2023. Among *de novo* PD cohort, ≤13 participants were missing a z-score for any one assessment; among HCs, 1 participant (not in the robust HC subgroup) had a missing z-score for Judgment of Line Orientation.

\* Cohen's d was computed by dividing the mean difference between groups by the pooled standard deviation.

Short title: Cognitive summary score in Parkinson's disease

**Table 4. Cognitive summary score using published and internally-derived norms**

Variable	PD Group	HC Group		Cohen's d*	
	All <i>de novo</i> PD (N = 923)	All HC (N = 250)	Robust HC (N = 154)	PD vs All HC	PD vs Robust HC
<b>Published norms</b>					
CSS, mean (SD)	-0.01 (0.64)	0.25 (0.61)	0.36 (0.61)	-0.42	-0.59
Median (min, max)	0.02 (-2.34, 1.99)	0.26 (-1.73, 1.99)	0.32 (-1.73, 1.99)		
<b>Internally-derived norms</b>					
CSS, mean (SD)	-0.40 (0.68)	-0.12 (0.62)	0.00 (0.61)	-0.41	-0.60
Median (min, max)	-0.34 (-2.75, 1.76)	-0.12 (-1.85, 1.83)	-0.03 (-1.85, 1.83)		

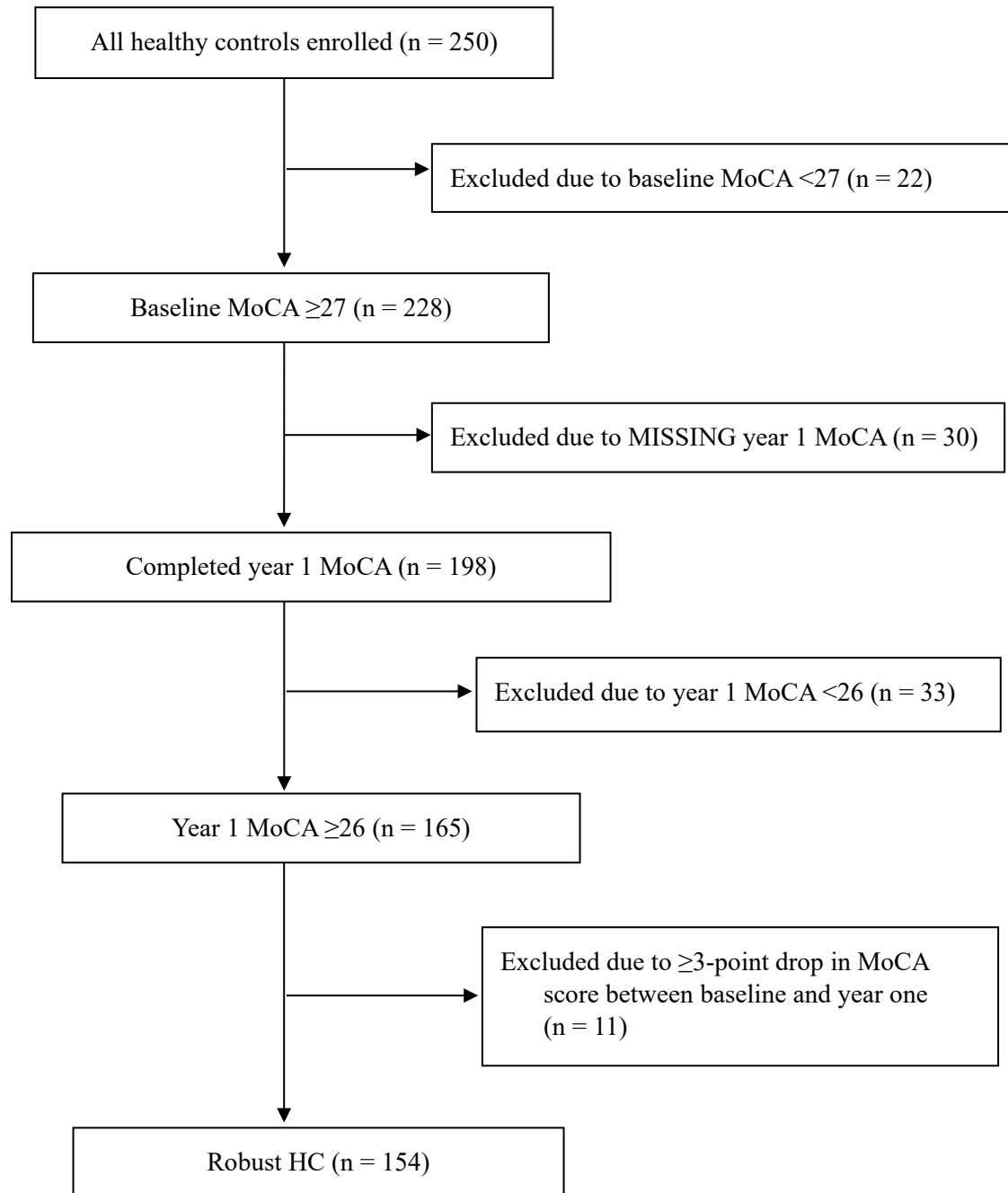
Abbreviations: CSS = cognitive summary score; HC = healthy controls; PD = Parkinson's disease.

Report Generated on Data Submitted as of: 06Nov2023. Among *de novo* PD cohort, 17 participants had a missing value for both cognitive summary scores; among HCs, 1 participant (not in the robust HC subgroup) had a missing value for both scores.

\* Cohen's d was computed by dividing the mean difference between groups by the pooled standard deviation.

Short title: Cognitive summary score in Parkinson's disease

**Figure 1. CONSORT Diagram for Super Healthy Controls**



Short title: Cognitive summary score in Parkinson's disease

**Supplementary Table 1. Regression models among robust healthy controls**

<b>Test</b>	<b>Variables</b>	<b><i>B</i></b>	<b>SE</b>	<b><i>t</i></b>	<b><i>p</i></b>	<b>Adj R<sup>2</sup></b>	<b>RMSE</b>
<b>HVLT-R Immediate Recall</b>	Intercept	28.37542	2.157	13.154	<.001	0.047	4.32146
	Age	-0.05979	0.030	-1.972	0.050		
	Sex*	1.58306	0.717	2.208	0.029		
<b>HVLT-R Delayed Recall</b>	Intercept	10.92487	1.060	10.309	<.001	0.070	2.12293
	Age	-0.03996	0.015	-2.682	0.008		
	Sex*	0.80139	0.352	2.275	0.024		
<b>Judgment of Line Orientation</b>	Intercept	15.07467	1.184	12.731	<.001	0.132	1.77892
	Age	-0.02374	0.012	-1.899	0.059		
	Sex*	-1.29422	0.297	-4.354	<.001		
	Education	0.08678	0.048	1.800	0.074		
<b>Letter Number Sequencing</b>	Intercept	12.30761	1.572	7.830	<.001	0.056	2.68988
	Age	-0.05359	0.019	-2.845	0.005		
	Education	0.13650	0.072	1.886	0.061		
<b>Semantic Fluency (Animal)</b>	Intercept	20.25284	1.338	15.142	<.001	0.016	5.49634
	Sex*	1.71525	0.908	1.889	0.061		
<b>Symbol Digit Modalities Test</b>	Intercept	76.80621	3.772	20.360	<.001	0.268	8.95934
	Age	-0.47289	0.063	-7.554	<.001		

Abbreviations: HC = healthy controls; HVLT-R = Hopkins Verbal Learning Test-Revised; PD = Parkinson's disease; RMSE = root mean square error.

Report Generated on Data Submitted as of: 06Nov2023.

\* Defined as 1=male and 2=female.

Short title: Cognitive summary score in Parkinson's disease

## REFERENCES

1. Aarsland D, Andersen K, Larsen J, Lolk A, Kragh-Sørensen P. Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Archives of Neurology*. 2003;60:387-392.
2. Hely M, Reid W, Adena M, Halliday G, Morris J. The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Movement Disorders*. 2008;23:837-844.
3. Aarsland D, Bronnick K, Williams-Gray C, et al. Mild cognitive impairment in Parkinson's disease: A multicenter pooled analysis. *Neurology*. 2010;75:1062-1069.
4. Weintraub D, Simuni T, Caspell-Garcia C, et al. Cognitive performance and neuropsychiatric symptoms in early, untreated Parkinson's disease. *Movement Disorders*. 2015;30(7):919-927.
5. Aarsland D, Bronnick K, Larsen J, Tysnes O, Alves G, Group FtPS. Cognitive impairment in incident, untreated Parkinson disease: The Norwegian ParkWest Study. *Neurology*. 2009;72:1121-1126.
6. Weintraub D, Aarsland D, Biundo R, Dobkin R, Goldman J, Lewis S. Management of psychiatric and cognitive complications in Parkinson's disease. *BMJ*. 2022;379:e068718.
7. Fantini ML, Farini E, Ortelli P, et al. Longitudinal study of cognitive function in idiopathic REM sleep behavior disorder. *Sleep*. 2011;34(5):619-625.
8. Chahine LM, Weintraub D, Hawkins KA, et al. Cognition in individuals at risk for Parkinson's: Parkinson associated risk syndrome (PARS) study findings. *Mov Disord*. 2016;31(1):86-94.
9. Litvan I, Goldman J, Troster A, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force Guidelines. *Movement Disorders*. 2012;27(3):349-356.
10. Litvan I, Kiebertz K, Troster AI, Aarsland D. Strengths and challenges in conducting clinical trials in Parkinson's disease mild cognitive impairment. *Mov Disord*. 2018;33(4):520-527.
11. Emre M, Aarsland D, Albanese A, et al. Rivastigmine for dementia associated with Parkinson's disease. *New England Journal of Medicine*. 2004;351:2509-2518.
12. Seppi K, Ray Chaudhuri K, Coelho M, et al. Update on treatments for nonmotor symptoms of Parkinson's disease-an evidence-based medicine review. *Mov Disord*. 2019;34:180-198.
13. Folstein M, Folstein S, McHugh P. "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975;12:189-198.
14. Nasreddine Z, Phillips N, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*. 2005;53:695-699.
15. Hoogland J, van Wanrooij LL, Boel JA, et al. Detecting mild cognitive deficits in Parkinson's disease: comparison of neuropsychological tests. *Mov Disord*. 2018;33:1750-1759.
16. Boel JA, de Bie RMA, Schmand BA, et al. Level I PD-MCI using global cognitive tests and the risk for Parkinson's disease dementia. *Mov Disord Clin Pract*. 2022;9(4):479-483.

Short title: Cognitive summary score in Parkinson's disease

17. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol.* 2014;71(8):961-970.
18. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. *J Neurol Neurosurg Psychiatry.* 2016;87(9):993-999.
19. Stout J, Queller S, Baker K, et al. HD-CAB: a cognitive assessment battery for clinical trials in Huntington's disease. *Movement Disorders.* 2014;29:1281-1288.
20. Weintraub D, Chahine L, Hawkins K, et al. Cognition and the course of prodromal Parkinson's disease. *Movement Disorders.* 2017;32:1640-1645.
21. Wyman-Chick KA, Martin PK, Weintraub D, et al. Selection of normative group affects rates of mild cognitive impairment in Parkinson's disease. *Mov Disord.* 2018;33(5):839-843.
22. Pedraza O, Lucas J, Smith G, Petersen R, Graff-Radford N, Ivnik R. Robust and expanded norms for the Dementia Rating Scale. *Archives of Clinical Neuropsychology.* 2010;25:347-358.
23. Marek K, Jennings D, Lasch S, et al. The Parkinson Progression Marker Initiative (PPMI). *Progress in Neurobiology.* 2011;95:629-635.
24. Marek K, Chowdhury S, Siderowf A, et al. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann Clin Transl Neurol.* 2018;5(12):1460-1477.
25. Holtzer R, Goldin Y, Zimmerman M, Katz M, Buschke H, Lipton RB. Robust norms for selected neuropsychological tests in older adults. *Arch Clin Neuropsychol.* 2008;23(5):531-541.
26. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's disease. *N Engl J Med.* 2023;388(1):9-21.
27. Papp KV, Rofael H, Veroff AE, et al. Sensitivity of the Preclinical Alzheimer's Cognitive Composite (PACC), PACC5, and Repeatable Battery for Neuropsychological Status (RBANS) to amyloid status in preclinical Alzheimer's disease -Atabecestat Phase 2b/3 EARLY Clinical Trial. *J Prev Alzheimers Dis.* 2022;9(2):255-261.
28. Sperling RA, Donohue MC, Raman R, et al. Trial of solanezumab in preclinical Alzheimer's disease. *N Engl J Med.* 2023;389(12):1096-1107.
29. Considine CM, Rossetti MA, Anderson K, et al. Huntington study group's neuropsychology working group position on best practice recommendations for the clinical neuropsychological evaluation of patients with Huntington disease. *Clin Neuropsychol.* 2024;38(4):984-1006.