Parkinsonism and Related Disorders 124 (2024) 107016

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Replication and reliability of Parkinson's disease clinical subtypes

Therese V. Cash [a], Christina N. Lessov-Schlaggar [b], Erin R. Foster [a,b,f], Peter S. Myers [a], Joshua J. Jackson [c], Baijayanta Maiti [a,d], Paul T. Kotzbauer [a], Joel S. Perlmutter [a,d,e,f,g], Meghan C. Campbell [a,d,*]

[a] Department of Neurology, Washington University School of Medicine, USA
[b] Department of Psychiatry, Washington University School of Medicine, USA
[c] Department of Psychological and Brain Sciences, Washington University in St. Louis, USA
[d] Department of Radiology, Washington University School of Medicine, USA
[e] Department of Neuroscience, Washington University School of Medicine, USA
[f] Program in Occupational Therapy, Washington University School of Medicine, USA
[g] Program in Physical Therapy, Washington University School of Medicine, USA

ARTICLE INFO

Keywords:
Parkinson disease
Latent class analysis
Classification
Psychometrics

ABSTRACT

Background: We recently identified three distinct Parkinson's disease subtypes: "motor only" (predominant motor deficits with intact cognition and psychiatric function); "psychiatric & motor" (prominent psychiatric symptoms and moderate motor deficits); "cognitive & motor" (cognitive and motor deficits).
Objective: We used an independent cohort to replicate and assess reliability of these Parkinson's disease subtypes.
Methods: We tested our original subtype classification with an independent cohort ($N = 100$) of Parkinson's disease participants without dementia and the same comprehensive evaluations assessing motor, cognitive, and psychiatric function. Next, we combined the original ($N = 162$) and replication ($N = 100$) datasets to test the classification model with the full combined dataset ($N = 262$). We also generated 10 random split-half samples of the combined dataset to establish the reliability of the subtype classifications. Latent class analyses were applied to the replication, combined, and split-half samples to determine subtype classification.
Results: First, LCA supported the three-class solution – Motor Only, Psychiatric & Motor, and Cognitive & Motor– in the replication sample. Next, using the larger, combined sample, LCA again supported the three subtype groups, with the emergence of a potential fourth group defined by more severe motor deficits. Finally, split-half analyses showed that the three-class model also had the best fit in 13/20 (65%) split-half samples; two-class and four-class solutions provided the best model fit in five (25%) and two (10%) split-half replications, respectively.
Conclusions: These results support the reproducibility and reliability of the Parkinson's disease behavioral subtypes of motor only, psychiatric & motor, and cognitive & motor groups.

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disease that includes heterogeneous patterns of motor, cognitive, and psychiatric dysfunction. Interest in PD subtypes emerged based on observed clinical differences in symptom profiles and patterns of progression [1]. Implications of establishing reliable and valid PD subtypes include improved prediction of clinical prognosis, increased accuracy of participant selection for clinical trials, and facilitated development of targeted interventions for specific features of PD [2]. Thus, reliable subtype classification may provide substantial clinical and research benefit.

For more than a decade, PD subtypes have been proposed to enhance our understanding of PD's heterogeneous clinical presentations. Yet, research on PD clinical subtypes has primarily focused on one cluster of symptoms at a time, such as the presence or severity of motor symptoms [3,4], global cognitive function [5,6], or psychiatric function [5,7,8]. Motor deficits have historically been the focus of PD subtyping, with tremor-dominant (TD) and postural instability and gait difficulty (PIGD) as the most identified motor subtypes [3]. However, these motor subtypes may better represent different stages of disease [4], fail to apply to a large number of cases, and have not been consistently replicated [9].

* Corresponding author.Washington University School of Medicine, Campus Box 8111, 660 South Euclid Avenue, St. Louis, MO, 63110, USA.
E-mail address: meghanc@wustl.edu (M.C. Campbell).

https://doi.org/10.1016/j.parkreldis.2024.107016
Received 12 December 2023; Received in revised form 24 April 2024; Accepted 19 May 2024
Available online 22 May 2024

Unfortunately, these early subtyping approaches often neglected to incorporate non-motor symptoms alongside motor profiles [1,10].

In addition, few studies have taken the important next steps to assess the reliability of the proposed subtype groupings [11,12]. Notably, one group used cluster analysis to classify PD participants into three subtype groups: motor/slow progression, diffuse/malignant, and intermediate. They also showed that the malignant phenotype was associated with more rapid motor, cognitive, and non-motor progression [13]. Subsequently, they successfully replicated the original groupings in a larger, independent sample [14]. A different group identified four subtype clusters defined by 1) fast motor progression with symmetrical motor disease; 2) mild motor and non-motor disease with intermediate motor progression; 3) severe motor disease with poor psychological well-being and intermediate motor progression; and 4) slow motor progression with tremor dominant, unilateral disease [15]. These four subtype groups were replicated across two independent, early PD cohorts and shown to be associated with levodopa response and motor progression rates [15]. However, another recent large scale study failed to reproduce eight data-driven PD subtype classifications, including subtypes based on age at onset, dominance of tremor, disease progression, non-motor symptoms, and presence of cognitive impairment [16]. A recent review of PD subtyping points out that methodological differences such as sample size and characteristics, use of a priori hypotheses vs. data-driven analytic approaches, and domains assessed could account for these mixed results [12]. Thus, additional subtype replication studies are needed to continue to build consensus about the most clinically relevant and reliable subtype classifications across the PD disease trajectory, including among more heterogenous patient samples who are taking medication.

In a non-demented heterogenous sample of treated PD participants, we recently identified three distinct clinical behavioral subtypes of PD [17]: “Motor Only”, with mild motor deficits; “Psychiatric & Motor”, with increased psychiatric symptoms; and “Cognitive & Motor”, with decreased cognitive function. Subsequently, these PD subtypes predicted mortality and dementia rates [17], lending support for their prognostic validity. Further, longitudinal analyses of baseline subtype classification groups indicated relatively greater decline in class-defining symptoms. Thus, the Motor Only class did not show cognitive or psychiatric problems, the Psychiatric & Motor class continued to have more prominent psychiatric symptomatology compared to the other classes, and the Cognitive & Motor class had greater decrease in cognitive function compared to the other classes [18]. Taken together, these findings show promise for a more comprehensive approach to PD subtype classification driven by psychiatric and cognitive symptoms. However, the reliability of these subtypes requires more evidence to ensure utility for further investigation, including biomarker and clinical translation efforts.

This study aims to replicate and assess reliability of the three clinical behavioral subtypes—“Motor Only”, “Psychiatric & Motor” and “Cognitive & Motor”— using an independent replication sample, combined sample, and split-half reliability samples. First, we hypothesized that the three PD subtypes would be recapitulated using an independent sample, providing justification for combining the original and replication samples for further analyses. Second, we hypothesized that the three PD subtypes would be reliably reproduced using a larger sample combining the original and replication samples. Finally, we predicted that the three-class solution would provide the best model fit in ten split-half samples, further supporting the internal reliability of the subtypes.

2. METHODS

2.1. Study protocol approvals, registrations, and patient consents

The Human Research Protection Office at Washington University in St. Louis approved this study. All participants provided written informed consent to participate either in the Protein and Imaging Biomarkers study (PIB) or the Protein Aggregation and Neurotransmitter Deficits (PAND) study at Washington University in St. Louis School of Medicine.

2.2. Participants

All participants come from two larger, longitudinal studies [17,18] examining Parkinson disease progression and healthy aging. For inclusion in the larger studies, all participants needed to be at least 50 years old for PIB and 55 years old for PAND, have a minimum of 12 years of education, and agree to brain donation. Healthy control participants from the larger study were not included in the original subtyping analyses or in the current replication sample analyses. PD participants had a clinical diagnosis of PD based on the modified United Kingdom PD Society Brain Bank clinical diagnostic criteria [19], with clear motor response to levodopa. For the larger studies, the exclusion criteria included: 1) other neurologic diagnoses, 2) head injury with loss of consciousness >5 min or neurologic sequelae, and 3) schizophrenia or bipolar disorder.

The original subtyping analysis [17] included 162 PD participants from the PIB study who completed baseline evaluations between the years 2006 and 2015. The replication sample included 87 PAND participants and 13 PIB participants (total $N = 100$), not previously included in the original sample, who completed baseline evaluations between the years 2016 and 2020. As with the original sample, participants with dementia, as indicated by Clinical Dementia Rating (CDR) Scale [20] global score $\geq$1, Mini Mental Status Exam (MMSE) total <24, or clinician judgment, were excluded.

2.3. Data collection

All participants completed an initial baseline evaluation including comprehensive motor, cognitive, and psychiatric assessments [17,18]. Motor testing with the Unified Parkinson Disease Rating Scale motor subscale III (UPDRS3-Total) was performed by a movement disorder specialist at the study visit or from video recording [21]. UPDRS3-Total rating was broken down into bradykinesia, rigidity, tremor, and postural instability and gait disturbance (PIGD) sub-scores. Cognitive evaluations consisted of a comprehensive battery of neuropsychological tests for each cognitive domain consistent with Level-II criteria for MCI assessment [22] (see Supplemental Table 1 for full cognitive test battery and references). PD participants completed motor and cognitive assessments after overnight withdrawal of PD medications in the practically defined ‘OFF’ state. In addition, trained raters administered the Clinical Dementia Rating evaluation (CDR) [20] with a collateral source and PD participant to assess functioning across domains of memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. CDR interviews were performed with participants ‘ON’ medications. The CDR global score was calculated, with the following clinical cutoffs, CDR $\geq$1 = dementia; CDR = 0.5, cognitive impairment; CDR = 0, intact cognition. Psychiatric assessments included self-reported depression (Geriatric Depression Scale [GDS] [23]) and apathy (Frontal Systems Behavior Scale- Apathy subscale [FrSBe-A] [24]). The participant’s collateral source also completed the Neuropsychiatric Inventory Questionnaire (NPIQ) [25] as an overall measure of psychiatric symptoms and severity. Psychiatric measures were administered with participants ‘ON’ medications since they assess functioning over a period of time rather than a specific time point.

2.4. Data processing

As with the original sample, age, sex, and education-adjusted scaled scores for the cognitive tests and the FrSBE-A were computed based on test manuals and published normative data. Scaled scores were converted to z-scores and averaged within each domain. The motor subscales and depression ratings (GDS) were normalized within sample. In total, 11 indicator variables were used in the LCA models to identify

Table 1
Baseline Demographics and Clinical Characteristics for Original and Replication Samples. Values represent mean (SD), except sex and CDR reported as percentage or total number. NPIQ, Neuropsychiatric Inventory Questionnaire; MMSE, Mini Mental Status Exam; CDR, Clinical Dementia Rating scale. Significant differences between the two samples ($p < 0.05$) are marked in bold based on Independent Samples *t*-test for Age, Independent Samples Mann-Whitney U Test for all non-normally distributed continuous variables (Education, PD Onset, PD Duration, NPIQ, MMSE), and Chi-square test for categorical variables (Sex, CDR).

	Original Sample	Replication Sample	Significance test-p value
	$N = 162$	$N = 100$	
Age (yrs)	66.01 (7.67)	68.90 (8.48)	**0.005**
Sex (% Male)	61.7 %	57.0 %	0.264
Education (yrs)	16.00 (2.51)	15.85 (2.40)	0.514
PD Onset (age)	59.80 (8.58)	62.33 (9.17)	0.076
PD Duration (yrs)	6.22 (3.76)	4.89 (3.96)	**0.001**
NPIQ Total	2.67 (2.93)	2.97 (3.21)	0.332
MMSE	28.26 (1.54)	27.89 (1.67)	0.080
CDR (0/.5)	91/71	55/45	0.520

subtypes (Motor: bradykinesia, tremor, rigidity, PIGD; Cognitive: attention, memory, language, visuospatial and executive function; Psychiatric: depression, apathy), consistent with the original work done on these subtypes [17].

2.5. Baseline sample characteristics and differences

Table 1 shows baseline demographic and clinical characteristics for the original and replication samples. Samples were compared using independent samples *t*-test for normally distributed variables (age), Mann-Whitney *U* Test for non-normally distributed variables (age of PD motor onset), and chi-square tests for categorical variables (sex, CDR Global Score). The original sample was significantly younger at study enrollment and had a longer duration of PD motor symptoms than the replication sample (both *ps* < 0.05; Table 1). No other significant differences were found between the two samples. Independent sample *t*-tests on the 11 indicator variable z-scores showed significantly higher visuospatial cognitive z-scores in the original than the replication sample and sample equivalence on the remaining variables (Supplemental Table 2). Thus, the original and replication sample were nearly equivalent on key variables, which supports use of the replication and combined samples to demonstrate reproducibility and reliability of the subtypes.

2.6. Statistical analysis

2.6.1. Latent class analysis

Latent class analysis (LCA) helps identify unobservable or latent subgroups of individuals in a sample based on the individuals' pattern of response or measurements across multiple observed variables, referred to as indicator variables. Individuals with similar response/measurement patterns are grouped together into latent classes. Analysis proceeds by modeling 1, 2, 3, and up to *k* number of classes. The best fitting model is selected based on several criteria including 1) the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) where lower relative scores indicate better model fit; 2) the Lo-Mendell-Rubin Likelihood Ratio Test (LMR-LRT) of model parsimony which evaluates the fit of the model with *k* classes compared to the fit of the model with one fewer, *k-1,* classes where non-significant LMR-LRT p value ($p > .05$) suggests greater parsimony of the *k-1* model; 3) model entropy, an indicator of classification confidence ranging 0 to 1 where higher values indicate greater classification certainty. For each participant, posterior probabilities of latent class membership are estimated for each class and probabilities across classes sum to 1. Individuals are assigned to the class for which they have the highest posterior probability value. A value ≥ 0.7 indicates reliable individual assignment to a given class [26]. Ideally, individuals have high posterior probabilities for the class to which they are assigned and low posterior probabilities for the other classes, resulting in high overall model entropy; and 4) meaningful class distinctions and enough individuals per class to allow further statistical analysis.

An extensive analysis of the best model fit criteria to use to decide the number of latent classes in mixture models concluded that the BIC performs best across different latent variable models, followed by the

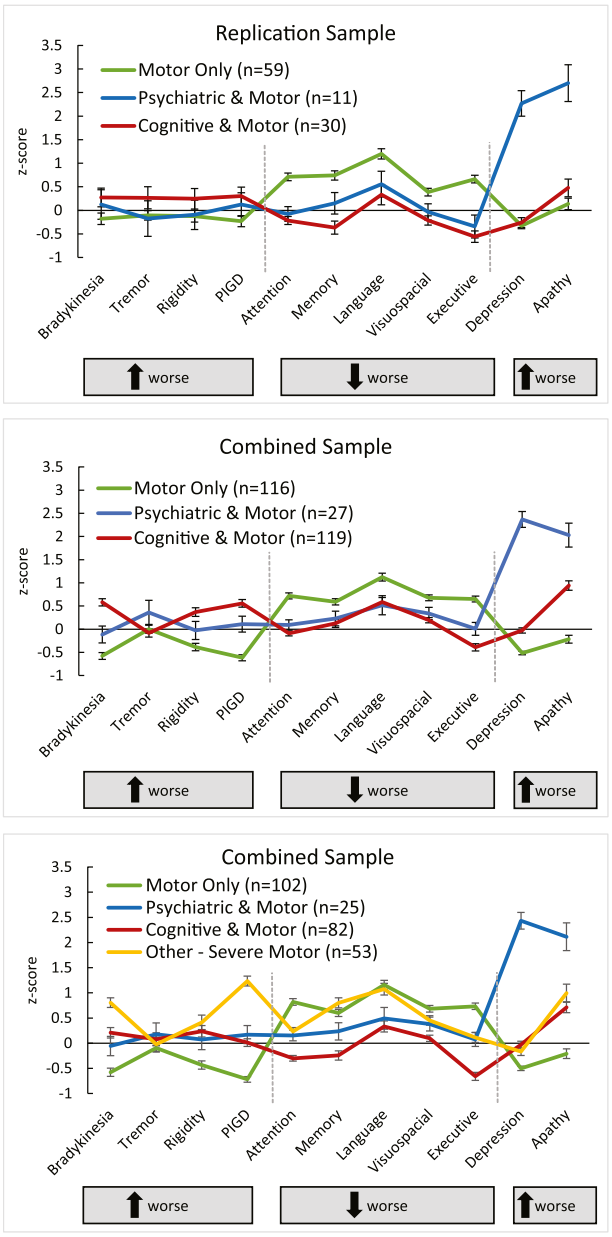

Fig. 1. Subtype differences across motor, cognitive, and psychiatric measures. Higher scores represent worse function for motor and psychiatric measures; lower scores represent worse function for cognitive domains. PIGD = postural instability and gait difficulty. **1a**) 3-class solution in replication sample ($N = 100$). Significant subtype differences ($p < .05$) for all measures, except bradykinesia ($p = .12$), tremor ($p = .22$) and rigidity ($p = .29$).Values represent z-scores for each measure (indicator). **1b**) 3-class solution replicated in combined sample ($N = 262$). Significant subtype differences ($p < .001$) for all measures, except tremor ($p = .13$). **1c**) 4-class solution emerged in combined sample ($N = 262$). Significant subtype differences ($p < .001$) for all measures, except tremor ($p = .54$).

Table 2
LCA model fit statistics for Replication and Combined Samples. Analysis proceeds by modeling 1, 2, 3, and up to k number of classes. The best fitting model is selected based on several model fit criteria including 1) the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) where lower relative scores indicate better model fit; 2) the Lo-Mendell-Rubin Likelihood Ratio Test (LMR-LRT) of model parsimony which evaluates the fit of the model with *k* classes compared to the fit of the model with one fewer, *k-1,* classes where non-significant LMR-LRT p value ($p > .05$) suggests greater parsimony of the *k-1* model; 3) Model entropy, a measure of classification uncertainty ranging 0 to 1 where higher values indicate greater classification certainty.

Sample	n	Classes	df	AIC	BIC	Adjusted BIC	Entropy	LMR-LRT p
Replication	100	1	22	2926.046	2983.359	2913.878		
		2	37	2840.321	2936.712	2819.857	0.966	0.4087
		3	**52**	**2778.882**	**2914.350**	**2750.121**	**0.961**	**0.4464**
		4	67	2750.102	2924.649	2713.046	0.878	0.3537
Combined	262	1	22	7609.060	7687.564	7617.814		
		2	37	7326.510	7458.538	7341.232	0.759	0.0149
		3	**52**	**7189.731**	**7375.285**	**7210.421**	**0.846**	**0.1022**
		4	67	7114.920	7353.999	7141.579	0.845	0.0784
		5	82	7061.928	7354.532	7094.555	0.816	0.5630

likelihood ratio test which was shown to select between neighboring models with high degree of confidence [24]. Based on these findings, our model selection placed greatest emphasis on the BIC, followed by the LMR-LRT test, and entropy last. We additionally considered interpretability and class prevalence to allow for further analyses using the classes as model variables. LCA was conducted using MPlus software (for reference, see Supplemental Table 1).

We investigated the reproducibility and internal reliability of the three-class solution by performing LCA on the independent replication sample ($N = 100$), on the combined original plus replication sample ($N = 262$), and on 10 split-half subsamples of the combined sample. Split-halves were generated using the following process: we generated 10 sets of 262 random numbers using the random number generator (RAND) function in Excel. The combined sample was sorted in ascending order for each of the 10 random number sets and each set was split into top and bottom halves resulting in a total of 20 replications. We controlled for age, sex, and education (number of years) in all models.

2.6.2. Discriminant analyses

Linear discriminant analyses were performed to determine how well the indicator variables discriminate between subtypes. Next, stepwise discriminant analyses were used to identify the key distinguishing features of the subtypes and identify the minimum set of variables that discriminate the subtype classifications. Discriminant analyses were performed on the replication sample ($N = 100$) and then on the combined sample ($N = 262$), to assess whether there were any differences between samples in discriminability or key features. For each discriminant analysis, classification results were reviewed to determine correct classification percentage based on the discriminant functions. In addition, stepwise discriminant analyses provided a list of the indicator variables providing significant predictive value to the discriminant function at each step (based on minimizing the Wilks' Lambda value at each step). Linear discriminant function analyses were conducted in SPSS (for reference, see Supplemental Table 1).

3. Results

3.1. Replication of PD clinical behavioral subtypes

LCA results support the reproducibility of the three clinical behavioral subtypes. Three distinct PD clinical subtypes were evident in the replication sample ($N = 100$), consistent with our findings in the original sample [17]. The three-class solution (see Fig. 1a) captures individuals with motor deficits (Motor Only, $n = 59$), individuals with motor deficits and high scores for depression and apathy (Psychiatric & Motor, $n = 11$), and individuals with motor deficits and low scores on cognitive measures (Cognitive & Motor, $n = 30$) (see Supplemental Table 3 for indicator variable values across subtypes).

The three-class solution had the lowest BIC and relatively high entropy (see Table 2). The LMR-LRT *p*-values were not informative with respect to model selection. Posterior probabilities for class assignments were high across classes: Motor Only ($M = .95$, *range* = 0.54–1.00), Psychiatric & Motor ($M = 0.98$, *range* = 0.89–1.00) and Cognitive & Motor ($M = 0.92$, *range* = 0.61–1.00). Just five out of 100 participants had a posterior probability below 0.70 for their assigned class ($n = 5/100$ or 5%), indicating low likelihood of misclassification.

3.2. Reliability of the clinical behavioral subtypes

LCA results for the combined ($N = 262$) and split-half samples ($N = 131 \times 20$) demonstrate reliability of the three clinical behavioral subtypes and the emergence of a possible fourth subtype. Model fit statistics for the combined sample are shown in Table 2. The 4-class solution had the lowest BIC, and the LMR-LRT *p* values suggested the three-class solution as best fit. LMR-LRT was not informative in the selection of the three- versus the four-class solution. It is possible that the combined larger sample size allowed for detection of an additional fourth class, therefore we show the results for both the three- and four-class solutions.

Fig. 1b shows the three-class solution using the combined data ($N = 262$), which captured the same Motor Only ($n = 116$), Psychiatric & Motor ($n = 27$), and Cognitive & Motor ($n = 119$) subgroups (see Supplemental Table 3 for indicator variables across subtypes). Fig. 1c shows the four-class solution, which included the same three classes plus a fourth class with more severely impaired motor function and intact cognition (Severe Motor). For the three-class solution, posterior probabilities remained high for all classes (Motor Only ($M = 0.94$, *range* = 0.54–1.00), Psychiatric & Motor ($M = 0.92$, *range* = 0.36–1.00), and Cognitive & Motor ($M = 0.93$, *range* = 0.50–1.00), with few participants exhibiting classification probabilities less than 0.70 ($n = 20/262$ or 7.6%). Again, this indicates a low likelihood for potential misclassification. Similarly, the four-class solution had high posterior probabilities for all groups (>0.89) and few participants with classification probabilities less than 0.70 ($n = 21/262$ or 8%). In the split-half samples, the three-class solution was the best fit model in 13 of 20 samples (65%), the two-class solution was the best fit model in five of 20 samples (25%), and the four-class solution was the best fit model in two of 20 samples (10%).

Notably, in the two split-half samples where a four-class solution was selected, similar to the results in the combined sample, the original three subtypes were still seen with the addition of a fourth class defined by more severe motor deficits but intact cognition and psychiatric functioning (Severe Motor). Thus, 75 % of the split-half samples supported the presence of the original three subtypes, with the potential identification of a fourth group.

3.3. Subtype discriminant features

Consistent with the original sample, discriminant analyses indicated robust subtype classification for the replication and combined samples.

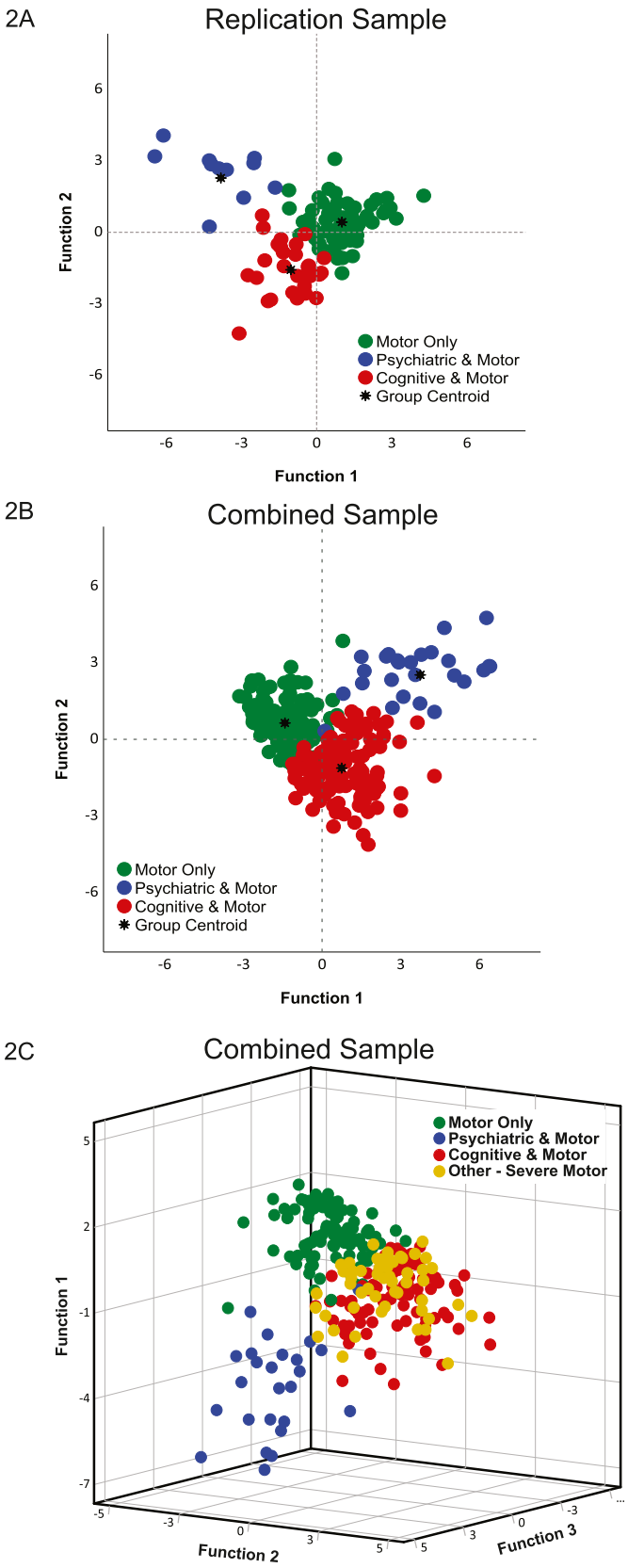

Fig. 2. Scatter plots for discriminant analyses of PD subtypes. **2a**) 3-class PD subtypes in replication sample ($N = 100$). Significant subtype separation with 95 % classification accuracy based on discriminant functions 1 and 2, which accounted for 63.0 % and 37.0 % of the variance, respectively. **2b**) 3-class subtypes in combined sample ($N = 262$). Significant subtype separation with 94 % classification accuracy based on discriminant functions 1 and 2, which accounted for 64.5 % and 35.5 % of the variance, respectively. **2c**) 4-class subtypes in combined sample ($N = 262$). Significant subtype separation with 93 % classification accuracy based on discriminant functions 1, 2, and 3, which accounted for 49.6 %, 32.5 %, and 17.9 % of the variance, respectively.

Fig. 2a shows the three-class solution in the replication sample yielding a 95% subtype classification rate. Fig. 2b also supports a robust three-class solution in the combined sample with a 94% rate of subtype classification. Fig. 2c shows the emergence of a four-class structure in the combined sample with a similarly high classification rate of 93%.

Stepwise discriminant analyses demonstrated that psychiatric and cognitive symptoms were key to the subtype classification models. To explore the contribution of each indicator variable to the overall model, stepwise discriminant analyses were performed on the three-class solution in the replication sample and the three-class and four-class solutions in the combined sample. In the replication sample, depression, executive function, memory, attention, and language best discriminated PD subtypes with 93.0% correct classification into the three subtype groups. In the combined sample, eight of the 11 indicator variables (depression, apathy, executive function, attention, language, PIGD, bradykinesia, and tremor) provided significant predictive value, with an overall correct classification rate of 92.7% for the three-class solution. With a four-class solution, a slightly different set of eight indicator variables (depression, apathy, executive function, attention, memory, language, PIGD and bradykinesia) significantly discriminated the subtype groups, with 93.1% classification rate.

4. Discussion

4.1. Summary

We investigated the reproducibility and reliability of three PD clinical subtypes defined by motor, psychiatric, and cognitive characteristics. Results supported the replication of the original PD clinical subtypes classified as Motor Only, Psychiatric & Motor, and Cognitive & Motor. LCA performed in a replication sample yielded the same three-class solution found in the original sample. Discriminant analyses indicated that similar clinical characteristics, primarily cognitive and psychiatric features, define the three subtypes across samples.

We also examined the reliability of the three PD clinical subtypes in a combined sample and in split-half samples. Across the combined and split-half samples, the three-class subtype solution was reliably reproduced. In the larger combined sample, a fourth class—defined by more severe motor symptoms— emerged. Importantly, the original three subtypes persisted as originally defined. Thus, the larger size of the combined sample allowed for expansion and potential refinement of the classification of the PD clinical subtypes, while retaining the original three subtypes. The split-half reliability analyses yielded the original three-class solution as the best fit, lending further support for the robustness of the original subtypes. In sum, the three-class subtype grouping, characterized by motor, cognitive, and psychiatric profiles, remains the most consistent across analyses.

This study adds to a growing literature focused on the reproducibility and application of PD subtypes to better understand the clinical and biological trajectory of PD. Similar to earlier clinical subtype work [13–15], our findings support a more comprehensive approach to PD clinical subtypes, incorporating motor and non-motor aspects to strengthen classification accuracy across patient cohorts. Of course, these clinical behavioral subtypes must be considered alongside emerging biological subtypes. Greater understanding of the neurobiology underlying the clinical features of PD and related syndromes holds enormous potential for improved diagnosis and treatment. Newly proposed staging systems of PD [27,28] offer a biological framework for a continuum of PD pathology from mild prodromal features to established disease. Our clinical behavioral subtypes are ripe for exploration as variants of disease expression that may inform or map onto the trajectory of biological progression. Anchoring biological subtypes to the heterogeneous clinical features of PD will be essential to maintain clinical relevance and improve quality of life for individuals with PD.

4.2. Strengths & limitations

This study is bolstered by using data-driven LCA methodology to examine the reproducibility and reliability of PD clinical subtypes within a replication sample, combined sample, and split-half samples. An advantage of data-driven LCA over a hypothesis-driven confirmatory approach is that analysis and interpretation are less prone to experimenter bias in model selection. Establishing reliability using both a combined sample as well as randomly generated split-half samples is another strength.

Limitations of this study include that the cohorts were recruited from a single site, and the same battery of assessment measures was used across the two samples. Thus, it remains possible that this subtype classification system would not hold up as robustly across other sites or using alternate assessment tools. Although it will be beneficial to investigate these PD subtypes using other assessment tools in the future, the current results are strengthened by using a battery of assessment tools known for their validity and reliability for measuring motor and non-motor deficits in PD [29].

5. Conclusions

Ultimately, this study provides an important benchmark for established reproducibility and reliability of three clinical behavioral subtypes for PD. Now, we can have greater confidence in exploring how these clinical behavioral subtypes relate to the biomarkers and neurobiological trajectory of PD. For example, do PD clinical subtypes relate to genetic differences? Will autopsy confirmation and brain tissue analyses differ based on subtype classification? How do the subtypes differ in response to medications and other treatments? Longitudinal analyses using latent transition analysis will further elucidate the stability and prognostic validity of these PD subtypes. An important next step will be to refine subtype classification methods to shorten the assessment battery needed to accurately predict subtype classification. This will facilitate clinical translation and implementation. In sum, this study provides a strong foundation for further investigation of these clinical behavioral subtypes and their relevance to improve clinical assessment and treatment of PD.

Funding sources

The authors gratefully acknowledge funding support for this work was provided by grants from NINDS (NS097437, NS075321, NS097799) and NIH NCRR (UL1RR024992); American Parkinson Disease Association (APDA) Advanced Research Center for PD at WUSTL; Greater St. Louis Chapter of the APDA; Oertli Fund; Paula & Rodger Riney Fund; Barnes Jewish Hospital Foundation (BJHF) (Elliot Stein Family Fund & PD Research Fund).

CRediT authorship contribution statement

Therese V. Cash: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis. **Christina N. Lessov-Schlaggar:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Erin R. Foster:** Writing – review & editing, Supervision, Project administration, Investigation, Data curation, Conceptualization. **Peter S. Myers:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Joshua J. Jackson:** Writing – review & editing, Methodology, Conceptualization. **Baijayanta Maiti:** Writing – review & editing, Data curation. **Paul T. Kotzbauer:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization. **Joel S. Perlmutter:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization. **Meghan C. Campbell:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation,

Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Therese Cash received support from NIH (NS097437, NS075321, NS097799, NS124738, NS118146, AT011015). Christina N. Lessov-Schlaggar received support from NINDS NS097437. Erin R. Foster is funded by the National Institutes of Health (NIA R21AG063974, NIA R01AG065214, NIDDK R01DK126826, R01DK064832), the American Parkinson Disease Association (APDA) (ID# 971,949), and the APDA Advanced Parkinson Disease Research Center at Washington University in St Louis, and the Missouri Chapter of the APDA. Peter S. Myers received support from NINDS NS097437. Joshua J. Jackson received support from NINDS NS097437, NIMH AG061162-01, NIDCD DC017522-02S1. Baijayanta Maiti received funding from NINDS K23 NS125107, National Center for Advancing Translational Sciences of the National Institute of Health KL2 TR002346, American Academy of Neurology and American Brain Foundation Clinical Research Training Fellowship in Parkinson's disease, Parkinson Study Group/Parkinson's Disease Foundation Mentored Clinical Research Award, Missouri Chapter of American Parkinson Disease Association, and the Jo Oertli fund and has received honoraria for reviewing grants as a member of the Parkinson Study Group mentoring committee and from the American Academy of Neurology for authorship. Paul T. Kotzbauer received funding from National Institutes of Health NS110436, NS097799, NS075321, NS110456, AG071754, NS123860, NS097437, and from Biogen and has received previous funding and honoraria from AbbVie. Joel S. Perlmutter has received research funding from National Institutes of Health NS075321, NS103957, NS107281, NS092865, U10NS077384, NS097437, U54NS116025, U19 NS110456, AG050263, AG-64937, NS097799, NS075527, ES029524, NS109487, R61 AT010753, (NCATS, NINDS, NIA), R01NS118146, R01AG065214, Department of Defense (DOD W81XWH-217-1-0393), Michael J FoxFoundation, Barnes-Jewish Hospital Foundation (Elliot Stein Family Fund and Parkinson disease research fund), American Parkinson Disease Association (APDA) Advanced Research Center at Washington University, Missouri Chapter of the APDA, Paula and Rodger Riney Fund, Jo Oertli Fund, Huntington Disease Society of America, Murphy Fund, N. Grant Williams Fund and CHDI. He co-directs the Dystonia Coalition, which received the majority of its support through the NIH (grants NS116025, NS065701 from the National Institutes of Neurological Disorders and Stroke, TR 001456 from the Office of Rare Diseases Research at the National Center for Advancing Translational Sciences). Dr. Perlmutter has provided medical legal consultation to Wood, Cooper and Peterson, LLC and to Simmons and Simmons LLP. He serves as Director of Medical and Scientific Advisory Committee of the Dystonia Medical Research Foundation, Chair of the Scientific Advisory Committee of the Parkinson Study Group, Chair of the Standards Committee of the Huntington Study Group (honoraria for this one), member of the Scientific Advisory Board of the APDA, Chair of the Scientific and Publication Committee for ENROLL-HD (honoraria from this one), and member of the Education Committee of the Huntington Study Group (honoraria from this one). Dr. Perlmutter has received honoraria from CHDI, Huntington Disease Study Group, Parkinson Study Group, Beth Israel Hospital (Harvard group), U Pennsylvania, Stanford U. Boston University. Meghan C. Campbell receives research support from NIH (NS097437, NS075321, NS097799, NS124378, AG063974, AT010753, AT010753-S1), the McDonnell Center for Systems Neuroscience, the Mallinckrodt Institute of Radiology at WUSTL, the Neurimaging Labs Innovation Award and has received honoraria from the Parkinson Foundation.

Acknowledgements

We thank the study participants for their time and effort to aid our understanding of Parkinson disease. We also thank the study coordinators and research nurse coordinators at Washington University School of Medicine, Phil Lintzenich, Thomas Belcher, Jenny Zhen-Duan, Anja Pogarcic, My Vu, Jenny Petros, Barb Merz, Katharine Cummings, Selma Avdagic, Kelly McVey, Andrea Slavik, Chris Waller, Jake Wolf, and especially Johanna Hartlein, for assistance with data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.parkreldis.2024.107016.

References

[1] M.A. Thenganatt, J. Jankovic, Parkinson disease subtypes, JAMA Neurol. 71 (2014) 499–504.
[2] C. Marras, K.R. Chaudhuri, N. Titova, T.A. Mestre, Therapy of Parkinson's disease subtypes, Neurotherapeutics 17 (2020) 1366–1377.
[3] J. Jankovic, M. McDermott, J. Carter, S. Gauthier, C. Goetz, L. Golbe, S. Huber, W. Koller, C. Olanow, I. Shoulson, Variable expression of Parkinson's disease: a base-line analysis of the datatop cohort, The parkinson study group, Neurology 40 (1990) 1529–1534.
[4] J.G. Nutt, Motor subtype in Parkinson's disease: different disorders or different stages of disease? Mov. Disord. 7 (2016) 957–961.
[5] R. Erro, C. Vitale, M. Amboni, M. Picillo, M. Moccia, K. Longo, G. Santangelo, A. De Rosa, R. Allocca, F. Giordano, G. Orefice, G. De Michele, L. Santoro, M. T. Pellecchia, P. Barone, The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients, PLoS One 8 (2013) e70244.
[6] S.M. van Rooden, F. Colas, P. Martinez-Martin, M. Visser, D. Verbaan, J. Marinus, R.K. Chaudhuri, J.N. Kok, J.J. van Hilten, Clinical subtypes of Parkinson's disease, Mov. Disord. 26 (2011) 51–58.
[7] J.M. Graham, H.J. Sagar, A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes, Mov. Disord. 14 (1999) 10–20.
[8] T.D. van Balkom, C. Vriend, H.W. Berendse, E.M. Foncke, Y.D. van der Werf, O. A. van den Heuvel, M. Klein, Profiling cognitive and neuropsychiatric heterogeneity in Parkinson's disease, Parkinsonism Relat. Disorders 28 (2016) 130–136.
[9] T. Simuni, C. Caspell-Garcia, C. Coffey, S. Lasch, C. Tanner, K. Marek, How stable are parkinson's disease subtypes in de novo patients: Analysis of the ppmi cohort? Parkinsonism Relat. Disorders 28 (2016) 62–67.
[10] S. Shakya, J. Prevett, X. Hu, R. Xiao, Characterization of Parkinson's disease subtypes and related attributes, Front. Neurol. 13 (2022) 810038.
[11] S.M. van Rooden, W.J. Heiser, J.N. Kok, D. Verbaan, J.J. van Hilten, J. Marinus, The identification of Parkinson's disease subtypes using cluster analysis: a systematic review, Mov. Disord. 25 (2010) 969–978.
[12] T.A. Mestre, S.M. Fereshtehnejad, D. Berg, N.I. Bohnen, K. Dujardin, R. Erro, A. J. Espay, G. Halliday, J.J. van Hilten, M.T. Hu, B. Jeon, C. Klein, A.F.G. Leentjens, J. Marinus, B. Mollenhauer, R. Postuma, R. Rajalingam, M. Rodríguez-Violante, T. Simuni, D.J. Surmeier, D. Weintraub, M.P. McDermott, M. Lawton, C. Marras, Parkinson's disease subtypes: critical appraisal and recommendations, J. Parkinsons Dis. 11 (2021) 395–404.
[13] S.M. Fereshtehnejad, S.R. Romenets, J.B. Anang, V. Latreille, J.F. Gagnon, R. B. Postuma, New clinical subtypes of Parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes, JAMA Neurol. 72 (2015) 863–873.
[14] S.M. Fereshtehnejad, Y. Zeighami, A. Dagher, R.B. Postuma, Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression, Brain 140 (2017) 1959–1976.
[15] M. Lawton, Y. Ben-Shlomo, M.T. May, F. Baig, T.R. Barber, J.C. Klein, D.M. A. Swallow, N. Malek, K.A. Grosset, N. Bajaj, R.A. Barker, N. Williams, D.J. Burn, T. Foltynie, H.R. Morris, N.W. Wood, D.G. Grosset, M.T.M. Hu, Developing and validating Parkinson's disease subtypes and their motor and cognitive progression, J. Neurol. Neurosurg. Psychiatry 89 (2018) 1279–1287.
[16] T.A. Mestre, S. Eberly, C. Tanner, D. Grimes, A.E. Lang, D. Oakes, C. Marras, Reproducibility of data-driven Parkinson's disease subtypes for clinical research, Parkinsonism Relat. Disorders 56 (2018) 102–106.
[17] M.C. Campbell, P.S. Myers, A.J. Weigand, E.R. Foster, N.J. Cairns, J.J. Jackson, C. N. Lessov-Schlaggar, J.S. Perlmutter, Parkinson disease clinical subtypes: key features & clinical milestones, Ann Clin Transl Neurol 7 (2020) 1272–1283.
[18] P.S. Myers, J.J. Jackson, A.K. Clover, C.N. Lessov-Schlaggar, E.R. Foster, B. Maiti, J.S. Perlmutter, M.C. Campbell, Distinct progression patterns across Parkinson disease clinical subtypes, Ann Clin Transl Neurol 8 (2021) 1695–1708.
[19] A.J. Hughes, S.E. Daniel, L. Kilford, A.J. Lees, Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases, J. Neurol. Neurosurg. Psychiatr. 55 (1992) 181–184.
[20] J.C. Morris, The clinical dementia rating (cdr): current version and scoring rules, Neurology 43 (1993) 2412–2414.
[21] B.A. Racette, S.D. Tabbal, D. Jennings, L.M. Good, J.S. Perlmutter, B.A. Evanoff, A rapid method for mass screening for parkinsonism, Neurotoxicology 27 (2006) 357–361.

[22] I. Litvan, J.G. Goldman, A.I. Troster, B.A. Schmand, D. Weintraub, R.C. Petersen, B. Mollenhauer, C.H. Adler, K. Marder, C.H. Williams-Gray, D. Aarsland, J. Kulisevsky, M.C. Rodriguez-Oroz, D.J. Burn, R.A. Barker, M. Emre, Diagnostic criteria for mild cognitive impairment in Parkinson's disease: movement disorder society task force guidelines, Mov. Disord. 27 (2012) 349–356.
[23] J.A. Yesavage, T.L. Brink, T.L. Rose, O. Lum, V. Huang, M. Adey, V.O. Leirer, Development and validation of a geriatric depression screening scale: a preliminary report, J. Psychiatr. Res. 17 (1982) 37–49.
[24] J. Grace, J.C. Stout, P.F. Malloy, Assessing frontal lobe behavioral syndromes with the frontal lobe personality scale, Assessment 6 (1999) 269–284.
[25] J.L. Cummings, M. Mega, K. Gray, S. Rosenberg-Thompson, D.A. Carusi, J. Gornbein, The neuropsychiatric inventory: comprehensive assessment of psychopathology in dementia, Neurology 44 (1994) 2308–2314.
[26] D.S. Nagin, C.L. Odgers, Group-based trajectory modeling in clinical research, Annu. Rev. Clin. Psychol. 6 (2010) 109–138.
[27] L.M. Chahine, K. Merchant, A. Siderowf, T. Sherer, C. Tanner, K. Marek, T. Simuni, Proposal for a biologic staging system of Parkinson's disease, J. Parkinsons Dis. 13 (2023) 297–309.
[28] G.U. Höglinger, C.H. Adler, D. Berg, C. Klein, T.F. Outeiro, W. Poewe, R. Postuma, A.J. Stoessl, A.E. Lang, A biological classification of Parkinson's disease: the SnyNeurGe research diagnostic criteria, Lancet Neurol. 23 (2024) 191–204.
[29] D. Berg, R.B. Postuma, B. Bloem, P. Chan, B. Dubois, T. Gasser, C.G. Goetz, G. M. Halliday, J. Hardy, A.E. Lang, I. Litvan, K. Marek, J. Obeso, W. Oertel, C. W. Olanow, W. Poewe, M. Stern, G. Deuschl, Time to redefine pd? Introductory statement of the mds task force on the definition of Parkinson's disease, Mov. Disord. 29 (2014) 454–462.