



# Characterization of Parkinson's Disease Subtypes and Related Attributes

Shamatree Shakya<sup>1</sup>, Julia Prevett<sup>1</sup>, Xiao Hu<sup>2,3,4</sup> and Ran Xiao<sup>1\*</sup>

<sup>1</sup> School of Nursing, Duke University, Durham, NC, United States, <sup>2</sup> School of Nursing, Emory University, Atlanta, GA, United States, <sup>3</sup> Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, United States, <sup>4</sup> Department of Computer Science, College of Arts and Sciences, Emory University, Atlanta, GA, United States

## OPEN ACCESS

### Edited by:

Carmen Rodriguez-Blazquez,  
Instituto de Salud Carlos III  
(ISCIII), Spain

### Reviewed by:

David Brooks,  
Imperial College London,  
United Kingdom  
Leire Ambrosio,  
University of Southampton,  
United Kingdom

### \*Correspondence:

Ran Xiao  
ran.xiao@duke.edu

### Specialty section:

This article was submitted to  
Movement Disorders,  
a section of the journal  
Frontiers in Neurology

**Received:** 05 November 2021

**Accepted:** 04 May 2022

**Published:** 23 May 2022

### Citation:

Shakya S, Prevett J, Hu X and Xiao R  
(2022) Characterization of Parkinson's  
Disease Subtypes and Related  
Attributes. *Front. Neurol.* 13:810038.  
doi: 10.3389/fneur.2022.810038

Parkinson's disease is a progressive neurodegenerative disease with complex, heterogeneous motor and non-motor symptoms. The current evidence shows that there is still a marked heterogeneity in the subtyping of Parkinson's disease using both clinical and data-driven approaches. Another challenge posed in PD subtyping is the reproducibility of previously identified PD subtypes. These issues require additional results to confirm previous findings and help reconcile discrepancies, as well as establish a standardized application of cluster analysis to facilitate comparison and reproducibility of identified PD subtypes. Our study aimed to address this gap by investigating subtypes of Parkinson's disease using comprehensive clinical (motor and non-motor features) data retrieved from 408 *de novo* Parkinson's disease patients with the complete clinical data in the Parkinson's Progressive Marker Initiative database. A standardized k-means cluster analysis approach was developed by taking into consideration of common practice and recommendations from previous studies. All data analysis codes were made available online to promote data comparison and validation of reproducibility across research groups. We identified two distinct PD subtypes, termed the severe motor-non-motor subtype (SMNS) and the mild motor- non-motor subtype (MMNS). SMNS experienced symptom onset at an older age and manifested more intense motor and non-motor symptoms than MMNS, who experienced symptom onset at a younger age and manifested milder forms of Parkinson's symptoms. The SPECT imaging makers supported clinical findings such that the severe motor-non-motor subtype showed lower binding values than the mild motor- non-motor subtype, indicating more significant neural damage at the nigral pathway. In addition, SMNS and MMNS show distinct motor (ANCOVA test:  $F = 47.35, p < 0.001$ ) and cognitive functioning ( $F = 33.93, p < 0.001$ ) progression trends. Such contrast between SMNS and MMNS in both motor and cognitive functioning can be consistently observed up to 3 years following the baseline visit, demonstrating the potential prognostic value of identified PD subtypes.

**Keywords:** Parkinson's disease, symptoms, subtypes, cluster analysis, data driven approach

## INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disease with a wide range of motor and non-motor manifestations (1). Motor features such as tremor, bradykinesia, rigidity, and postural instability are hallmarks for clinical assessment and diagnosis of PD (2). However, a growing body of evidence shows that PD is highly heterogeneous, including complex motor and non-motor features (3); and changes in brain imaging and cerebrospinal fluid (CSF) biomarkers (4–6). The detection, management, and assessment of PD can be incomplete and erroneous when non-motor features, imaging, and biospecimen changes are not considered. Both missed diagnosis and false diagnosis can severely affect the quality of life since misdiagnosis can result in undertreatment; thus, increasing the chance of rapid disease progression. Similarly, inaccurate diagnosis and treatment can expose individuals to serious side effects due to potent therapeutic medications (7). Given the heterogeneous nature of PD, it marks great clinical significance and benefits to identify underlying PD subtypes, which can be transformational in designing prevention and treatment strategies (8). Identifying subtypes of PD with shared characteristics can promote the discovery of PD etiology through the understanding of the underlying pathophysiological mechanism (1). In addition, the identification of subtypes opens the door for early detection and management of PD and the development of tailored management strategies, which help address the specific needs of an individual patient and prevent over/under treatment mishaps (8, 9).

Traditionally, motor symptoms are used for PD subtyping, with tremor-dominant (TD) and postural instability and gait difficulty (PIGD) as the most consistently identified motor subtypes (10, 11). These seminal studies provide crucial evidence on the existence of distinct subtypes of PD. However, non-motor features are grossly overlooked in the subtyping process. The fact that the choice of variables is driven by a priori hypothesis with the empirical section of cut-offs also introduces substantial ambiguity in the subtyping results (12). Since the first adoption of data-driven cluster analysis in PD subtyping (13), it has become a more popular approach in later studies in identifying PD subtypes. Cluster analysis can take a wide collection of variables into consideration for PD subtypes bypassing the a priori selection process, and determine the optimal subtypes based on statistical methods.

A body of literature exists utilizing cluster analysis with both motor and non-motor variables aiming to achieve an objective PD subtyping (13–18). Graham et al. identify three distinct PD subtypes, including “motor only,” “motor and cognition,” and “rapid progression” (13). Erro et al. also identify three PD subtypes but with different characteristics, with the first group bearing the lowest motor and non-motor burden, and the other two groups with a similar motor burden but different non-motor involvement (18). In another study, three distinct subtypes are defined as “mild motor-predominant,” “diffuse malignant,” and “intermediate” (17). Despite the positive findings, there are still wide inconsistencies in proposed subtypes of PD. One review paper compared subtyping results from two studies using the

same database (17, 18) and found only 56% agreement on cluster memberships (4). A possible reason for these inconsistencies may be due to inherent limitations of data-driven subtyping of PD using hierarchical and non-hierarchical cluster analysis (6). The inconsistent subtypes can also stem from heterogeneity in data sources, study designs, patient populations (19), motor and non-motor data, and cluster analysis techniques used for identifying subtypes. The inclusion of PD patients on and off medications could be another source for inconsistencies in proposed PD subtypes (18); inclusion of PD patients on and off medications can also lead to erroneous PD subtyping. On top of inconsistent results, the poor reproducibility of reported PD subtypes poses another challenge in the field. Mestre et al. validated PD subtyping results from eight studies using the same database over a 10-year span by a panel of clinical experts following a modified Delphi consensus process (19). It was found that 7 out of 8 studies were “not very well” or “not at all” reproduced. Considering these issues in PD subtyping, multiple review papers outlined the need for standardized application of cluster analysis and additional results to reconcile the inconsistencies in PD subtyping (4, 12, 19).

This study aims to mitigate these issues by identifying PD subtypes in *de novo* PD patients using the comprehensive motor and non-motor data from the Parkinson's Progressive Markers Initiative (PPMI) (20). Following a common study pathway as summarized in (12), the present study starts off implementing data-driven k-means cluster analysis with a standardized approach using recommended practice in (6) to obtain objective PD subtyping. Next, the study conducts *post-hoc* analyses to describe motor, non-motor, CSF biomarkers and neuroimaging characteristics of PD subtypes. Lastly, the study evaluates the longitudinal progression of motor and global cognitive functioning of identified PD subtypes and the stability of these subtypes up to 3 years following the baseline visits. Our results are compared to previous studies to confirm past findings and to investigate discrepancies. Given that the data used in the study is publicly accessible, we also make our data analysis codes publicly available (<https://github.com/ranxiao/PDSubtyping>) to share the standardized cluster analysis approach, and maximally facilitate result comparison and reproducibility by other research groups.

## MATERIALS AND METHODS

### Study Data and Population

Data used in this study were obtained from the PPMI database (20). PPMI is an ongoing longitudinal observational study involving participants from multiple clinical sites (21) situated in 11 countries. PPMI contains a wide range of information on demographics, motor, and non-motor features; Single-photon emission computed tomography (SPECT-DaT) brain imaging, cerebrospinal fluid (CSF) biomarker, and genetic biomarkers from three groups, namely (a) untreated *de novo* PD, (b) healthy control, and (c) SWEDD (scans without evidence of dopaminergic deficit) participants (20). We only included untreated *de novo* PD patients at the baseline to identify PD subtypes for this study. Clinical motor assessments were collected

every 3 months during the first year and every 6-month interval after that. Assessments on cognitive and behavioral features were collected at every 12-month interval. Imaging data were collected from PD patients at visits in 12, 24, and 36 months (20). The details of the PPMI dataset and cohorts are available on the PPMI website [www.ppmi-info.org](http://www.ppmi-info.org). This study has received an exemption from the Institutional Review Board (IRB) of Duke University as it is using publicly available deidentified data from the PPMI.

The PPMI includes comprehensive information on motor and non-motor features from PD patients. Following data were used in our study:

- 1) Demographic characteristics: age of symptom onset and sex.
- 2) Motor features: International Parkinson's disease and Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III total score and subscores: rigidity (item 3), tremor (sum of items 15–18), bradykinesia (sum of item 2, 4–9, and 14), and axial scores (sum of items 1 and 9–13) (22). Greater scores in motor features indicate greater severity of motor symptoms.
- 3) Non-motor features: Neuropsychiatric features—Montreal Cognitive Assessment (MoCA), psychological features from MDS-UPDRS Part I fatigue (item 13), pain (item 9), apathy (item 5), hallucinations (item 2), Geriatric Depression Scale (GDS), State of Anxiety Inventory (STAI) score, REM sleep behavior disorder (RBD); and autonomic features—olfactory dysfunction using the University of Pennsylvania Smell Identification Test (UPSIT) score, autonomic dysfunction using the Scales for Outcomes in Parkinson's disease Autonomic (SCOPA-AUT) total score and its sub item scores: orofacial, gastrointestinal, urinary, cardiovascular, thermoregulatory, pupilometer, and sexual scores. Greater scores in non-motor features indicate greater severity except MoCA and UPSIT scores.

We used the above-mentioned motor, non-motor data, and age of onset to identify Parkinson's disease subtypes and compared clinical attributes (motor, non-motor, brain imaging, CSF biomarkers, and symptom progressions) across the identified subtypes.

## Brain Imaging Markers

PPMI collected dopaminergic imaging—single-photon emission computed tomography (SPECT) from the PPMI participants. SPECT with CAT tracer I-ioflupane uses tracers that bind with dopamine cells, and SPECT imaging is a gold standard test to assess dopaminergic denervation in several neurodegenerative diseases, including Parkinson's disease (23, 24). A summary of imaging markers (right caudate, left caudate, right putamen, left putamen, right striatum, left striatum, mean caudate measure, mean putamen measure, mean striatum measure, lowest caudate measure, highest caudate measure, lowest putamen measure, highest putamen measure, lowest striatum measure, highest striatum measure, asymmetry index-caudate, putamen, striatum, left count density ratio, right count density ratio, contralateral striatum, contralateral caudate, contralateral putamen, contralateral count density ratio, ipsilateral striatum,

ipsilateral caudate, ipsilateral putamen, ipsilateral count density ratio) was used to compare neurodegenerative imaging changes between subtypes at the baseline.

## CSF Biomarkers

PPMI includes comprehensive CSF biomarkers. Strong evidence underscores that CSF biomarkers can be sensitive indicators of cognitive decline (25), dopaminergic deficit (26), neurodegenerative changes (27), motor features (28), disease progression and neurofibrillary changes in PD (29). CSF biomarkers are found to be associated with motor progression, neuropsychiatric features, and cognitive decline in Parkinson's disease patients (30). We used a summary of CSF biomarkers (CSF A $\beta$ 1-42,  $\alpha$ -synuclein, t tau, p-tau, t-tau/A $\beta$  1-42 ratio, p-tau/A $\beta$  1-42 ratio, p-tau/t-tau ratio, t-tau/ $\alpha$ -synuclein ratio, p-tau/ $\alpha$ -synuclein ratio) to compare CSF biomarkers across subtypes. Correlational analyses were also conducted to find the association of identified significant factors with motor and non-motor scores, respectively.

## Longitudinal Assessments

We compared the rate of motor and cognitive progression across the identified PD subtypes using Hoehn and Yahr's and MoCA scores. Patients included in the cluster analysis but without any of these symptom data in follow-up visits are identified and removed from the longitudinal assessments. Hoehn and Yahr's scale is demonstrated to have greater clinicometric properties and is highly correlated with the motor function of MDS-UPDRS (31). Hoehn and Yahr's scale has been used as a successful indicator of Parkinson's disease progression in clinical and clinical trial studies (32). The scores were assessed at the baseline (Y0), 12 months (Y1), 24 months (Y2), and 36 months (Y3) as described below in *post-hoc* analyses. A larger Hoehn and Yahr's score is an indication of severe motor dysfunction, whereas a lower MoCA score is reflective of greater cognitive impairment.

## K-Means Cluster Analysis

A total of 683 participants were in PPMI datasets, including Parkinson's disease, healthy control, and SWEDD (Scans without evidence of dopaminergic deficit). Among them, 423 untreated *de novo* participants with PD diagnosis at the baseline met the eligibility criteria for this study. After eliminating participants with missing values, 408 participants (males: 269, females: 139) with complete data were included for k-means cluster analysis. The average age of participants was 59.6 years (age range: 25–83 years). Twenty-two variables representing motor and non-motor features of PD were retrieved from PPMI datasets for cluster analysis. The corresponding clinical scores were standardized using min-max normalization. We conducted a k-means cluster analysis to identify the PD subtypes. A list of cluster numbers ranging from 2 to 10 was evaluated. The optimal number of subtypes was identified based on the Calinski-Harabasz (C-H) pseudo-F value. The cluster analysis was repeated 100 times with random initialization to avoid final cluster membership derived from a local optimum. The maximal number of iterations was set at 100 to ensure ample iterations to arrive at an optimal clustering

solution within each cluster analysis. All analyses were carried out using Matlab (Mathworks Inc., Natick, MA).

## Post-hoc Analyses

The descriptive statistics (mean and standard deviation) were performed to summarize and compare the motor, non-motor features, SPECT-DaT brain imaging, and CSF biomarker attributes of two identified subtypes. Three *post-hoc* analyses were performed to (a) summarize motor and non-motor attributes, (b) SPECT-DaT brain imaging and CSF biomarkers, and (c) characterize motor and global cognitive progression rates, using Hoehn and Yahr and MoCA scores, respectively. The mean and standard deviation are calculated for each of the 22 baseline attributes and statistically compared between the two subtypes. The Hedges' *g* standardized mean difference (SMD) and its 95% confidence interval (*via* bootstrapping 3,000 times) are calculated to reveal the level of differences between subtypes. SMD is calculated by

$$SMD_i = \frac{m_{1i} - m_{2i}}{\delta_i}, i = 1, \dots, 22 \quad (1)$$

where  $m_{1i}$  and  $m_{2i}$  are the mean values of  $i^{th}$  baseline attribute in subtypes I and II, respectively.  $\delta_i$  is the pooled standard deviation from both subtypes. The SMD is calculated using the measures of effect size (MES) toolbox ver.1.5 (33). The mean and standard deviation are calculated for each of 28 imaging measures and 10 CSF biomarkers, and statistically compared between the two subtypes. A non-parametric Wilcoxon rank-sum test was performed to examine if statistically significant differences existed between the identified subtypes in baseline clinical features (motor and non-motor features), imaging, and CSF biomarkers. The significant level  $\alpha$  is set at 0.05, and multiple comparisons are adjusted by Bonferroni correction.

Visualizing cluster memberships is an important qualitative evaluation that can offer additional insights into derived PD subtypes. Due to the multivariate nature of input features, it's challenging to visualize the clusters in a two-dimensional format. While dimensional reduction techniques, such as principal component analysis (PCA), can be implemented to visualize the separation of different clusters (17), the subsequent interpretation is non-trivial contributed by the highly abstract nature of principal components. Like the global composite outcome proposed in (17), the present study establishes two new composite scores, the motor composite score, and the non-motor composite score. This is calculated by rescaling all motor and non-motor features (as described in the above section) into the range of (0, 1) through the min-max normalization and averaging the normalized scores of all components in each category, respectively. Since MoCA and UPSIT have reversed severity scores compared to other non-motor features, they are adjusted by subtracting the normalized scores from one before the averaging operation for calculating the non-motor composite score. For both composite scores, the higher value indicates a severe condition. A scatter plot is generated to depict the distribution of cluster memberships of each sample in a two-dimensional plane representing the motor and non-motor symptom severity.

The progression is evaluated based on four time points, baseline visit (Y0), 12-month follow-up (Y1), 24-month follow-up (Y2), and 36-month follow-up (Y3). Three progression rates are derived to capture progression at various temporal resolutions, including early progression (change of scores between Y1 and Y0), secondary progression (change of scores between Y3 and Y1) and long-term progression (change of scores between Y3 and Y0). In addition, a time plot of motor and non-motor symptom progression is presented to qualitatively evaluate the stability of PD subtypes. An analysis of covariance (ANCOVA) test is performed to investigate the effect of PD subtypes and different visits on the symptom progression, which is adjusted with age at symptom onset as a covariate in the statistical analysis. The significant level  $\alpha$  is set at 0.05.

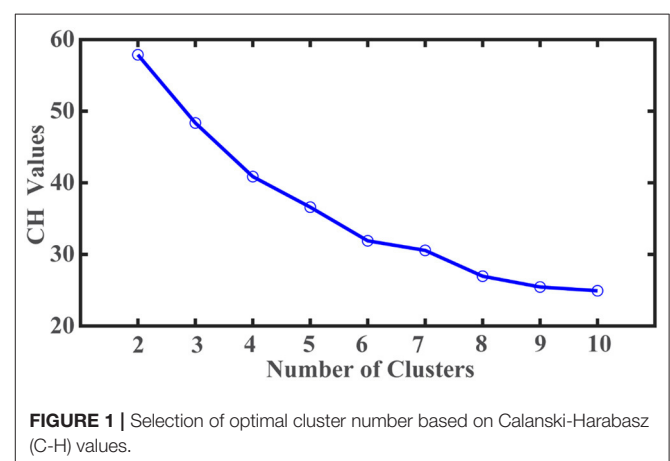
## RESULTS

### Baseline Description

A total of 408 *de novo* untreated PD patients comprising 269 (65.93%) males and 139 (34.96%) females with complete clustering data were included in the k-means cluster analysis.

### Statistical Clusters

There are twenty-two variables, including the age of symptom onset, motor, and non-motor features, used to identify PD subtypes based on the k-means cluster analysis. The clustering optimum was attained for the two-cluster solution. As shown in **Figure 1**, the largest C-H pseudo-F value obtained was 57 from the two-cluster solution. More samples are assigned to the PD subtype I ( $N = 270$ ) than the second subtype ( $N = 138$ ). PD subtype I includes participants who manifest symptoms at a younger age (young at symptoms onset) with an average age of  $58.2 \pm 10.2$  (age range = 25.4–80.1 years) than PD subtype II participants (older at symptoms onset) with an average age of  $62.4 \pm 8.9$  years (range = 35.6–83 years).



**TABLE 1** | Baseline clinical characteristics of PD subtypes.

Characteristics	Mild motor and non-motor subtype (MMNS, N = 270) Mean (SD)	Severe motor and non-motor subtype (SMNS, N = 138) Mean (SD)	p-value ( $\alpha = 0.0022$ )
Male sex (%)	177 (65.56%)	92 (66.67%)	-
Age of symptom onset *	58.177 ( $\pm 10.223$ )	62.363 ( $\pm 8.899$ )	<0.001
MDS UPDRS Part III*	17.989 ( $\pm 7.278$ )	26.167 ( $\pm 8.810$ )	<0.001
• Tremor	0.417 ( $\pm 0.314$ )	0.465 ( $\pm 0.314$ )	0.124
• Rigidity*	0.640 ( $\pm 0.460$ )	0.981 ( $\pm 0.583$ )	<0.001
• Bradykinesia*	0.708 ( $\pm 0.372$ )	1.078 ( $\pm 0.448$ )	<0.001
• Axial subscore*	0.248 ( $\pm 0.283$ )	0.481 ( $\pm 0.283$ )	<0.001
UPSIT*	23.830 ( $\pm 7.842$ )	19.268 ( $\pm 8.361$ )	<0.001
MoCA*	27.356 ( $\pm 2.299$ )	26.696 ( $\pm 2.316$ )	0.001
GDS*	1.682 ( $\pm 1.997$ )	3.601 ( $\pm 2.803$ )	<0.001
STAI*	61.226 ( $\pm 15.289$ )	73.341 ( $\pm 21.078$ )	<0.001
Hallucination	0.025 ( $\pm 0.159$ )	0.043 ( $\pm 0.204$ )	0.340
Apathy*	0.100 ( $\pm 0.313$ )	0.413 ( $\pm 0.681$ )	<0.001
Fatigue*	0.448 ( $\pm 0.587$ )	1.073 ( $\pm 0.965$ )	<0.001
Pain*	0.581 ( $\pm 0.755$ )	1.007 ( $\pm 0.956$ )	<0.001
RBD*	3.300 ( $\pm 2.053$ )	5.739 ( $\pm 2.960$ )	<0.001
SCOPA-AUT*	6.722 ( $\pm 3.639$ )	15.145 ( $\pm 6.387$ )	<0.001
SCOPA-gastrointestinal*	1.274 ( $\pm 1.352$ )	3.840 ( $\pm 2.169$ )	<0.001
SCOPA-urinary*	3.256 ( $\pm 2.148$ )	6.167 ( $\pm 3.536$ )	<0.001
SCOPA-cardiovascular*	0.300 ( $\pm 0.554$ )	0.789 ( $\pm 1.014$ )	<0.001
SCOPA-thermoregulatory*	0.800 ( $\pm 1.072$ )	1.935 ( $\pm 1.684$ )	<0.001
SCOPA-pupillomotor*	0.285 ( $\pm 0.575$ )	0.681 ( $\pm 0.715$ )	<0.001
SCOPA-sexual*	0.807 ( $\pm 1.264$ )	1.732 ( $\pm 1.878$ )	<0.001

\*Indicates statistical significance between the two subtypes (Wilcoxon rank-sum test).

## Variation in Baseline Clustering Characteristics

As shown in **Table 1**, we observed statistically significant differences in a majority of motor and non-motor features. There were statistically significant differences in age of symptom onset ( $p < 0.001$ ), MDS UPDRS III score ( $p < 0.001$ ), rigidity ( $p < 0.001$ ), bradykinesia ( $p < 0.001$ ), axial subscore ( $p < 0.001$ ), MoCA ( $p = 0.0014$ ), GDS ( $p < 0.001$ ), fatigue ( $p < 0.001$ ), pain ( $p < 0.001$ ), RBD ( $p < 0.001$ ), SCOPA-total ( $p < 0.001$ ), and SCOPA-subscores ( $p < 0.001$ ) between the two PD subtypes. No statistical significance was observed in their tremor ( $p = 0.1243$ ) and hallucination scores ( $p = 0.3408$ ). Overall, PD subtype II (older at symptom onset) experienced more severe Parkinson's symptoms (motor and non-motor symptoms) than the subtype I (younger at symptom onset). Subtype II experienced more severe cognitive impairment marked by a lower MoCA score than the subtype I. Based on the intensity of Parkinson's symptom burden, we termed the subtype I and the subtype II as "mild motor-non-motor subtype (MMNS)" and "severe motor-non-motor subtype

(SMNS)", respectively. Such observations are corroborated by the distribution of samples in the motor-non-motor plane, as presented in **Figure 2**. It shows SMNS samples (red dots) generally scatter on the top-right direction of MMNS samples (blue dots), with higher motor and non-motor composite scores. **Figure 3** reveals the level of differences between the two subtypes with SMDs and corresponding 95% CIs. It shows that SMNS demonstrates more severe symptoms at baseline than the MMNS regardless of motor and non-motor attributes. The majority of attributes have a large effect size between the two subtypes, with most absolute SMD values close to one.

## Variation in Imaging Markers

As shown in **Table 2**, SMNS and MMNS varied significantly in the brain imaging markers. Generally, the MMNS subtype demonstrated greater binding values in caudate, putamen, striatum regions than the SMNS subtype.

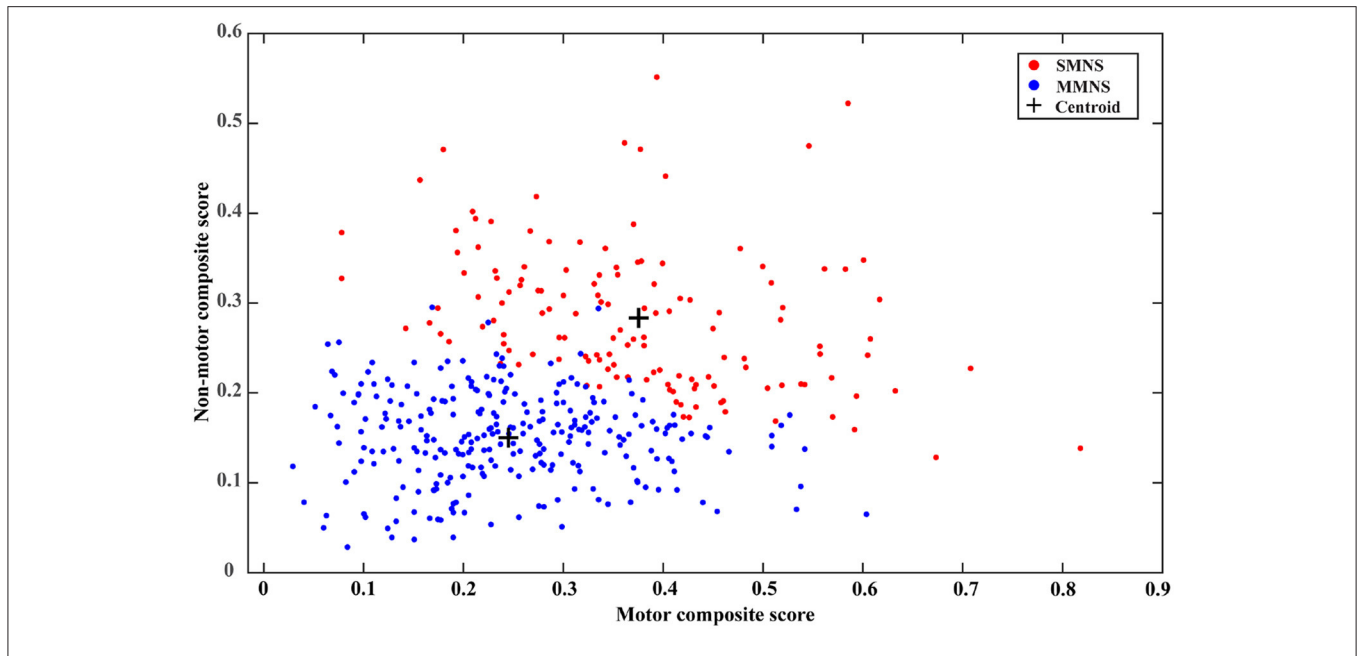
## Variation in CSF Biomarkers

The SMNS and MMNS subtypes did not vary significantly in terms of CSF biomarkers except the p-tau/ $\alpha$ -synuclein ratio ( $p = 0.002$ ), as shown in **Table 3**.

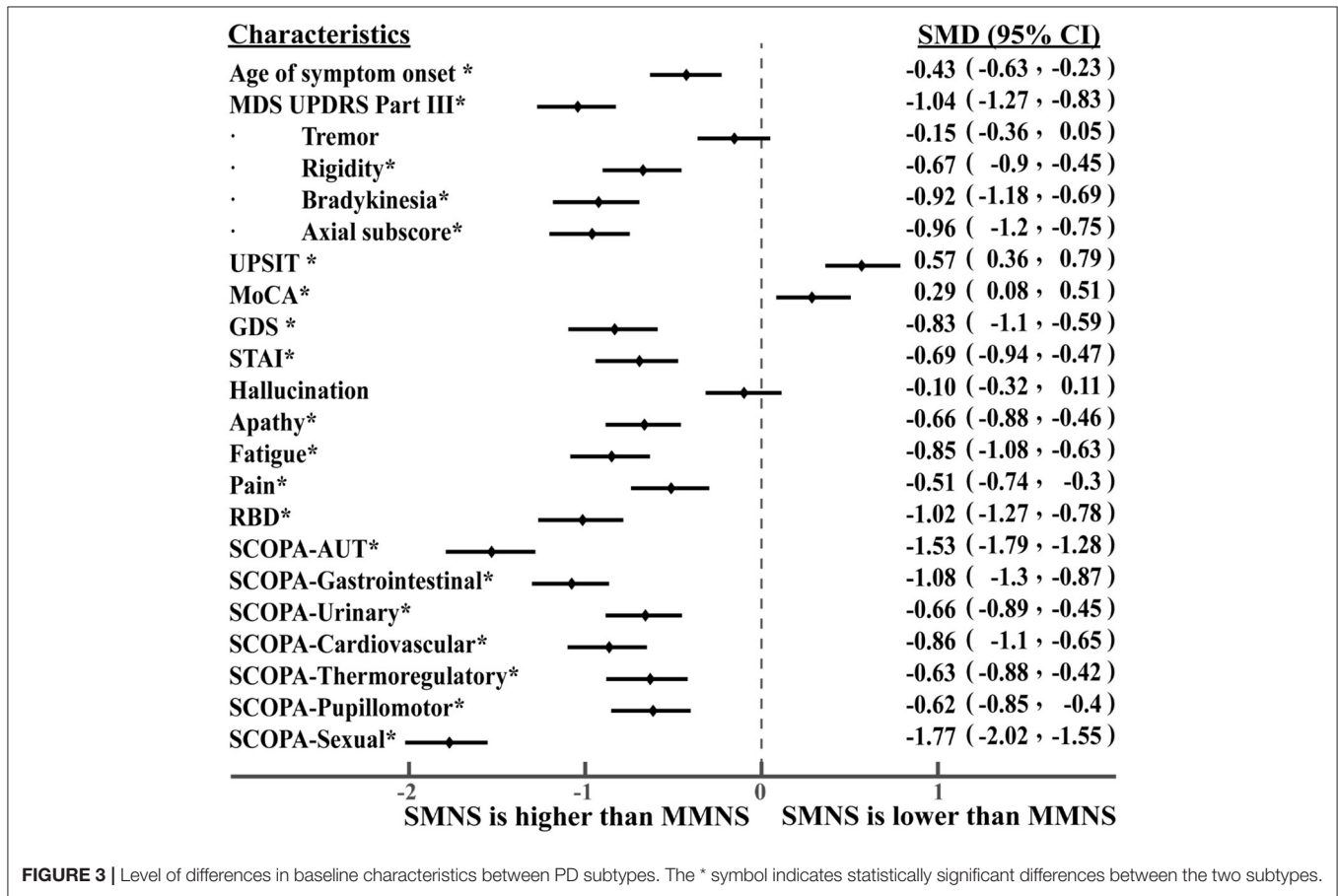
## Longitudinal Progression

There are 17 PD patients with baseline data but without symptom data in follow-up visits. This results in 391 patient data entering the longitudinal analysis with 262 in the MMNS group and 129 in the SMNS group, which maintain a similar patient ratio between two subtypes compared to those used in the cluster analysis. **Figure 4** depicts the time plot of H&Y (blue lines) and MoCA (red lines) scores from baseline up to 36-month follow-up. The MMNS subtype is presented in solid lines and the SMNS subtype in dash lines. For the progression of the motor symptom as captured in H&Y (blue lines), SMNS presents a more severe pattern than MMNS consistently across all four time points. SMNS also presents fewer variations in the changing rate compared to MMNS. As for non-motor symptoms (red lines), SMNS presents worse cognitive functions than MMNS consistently across all four time points. It also shows an increasing difference in cognitive functions along time between the two subtypes. The ANCOVA test shows there is a significant difference between the two subtypes in both motor ( $F = 47.35$ ,  $p < 0.001$ ) and non-motor ( $F = 27.67$ ,  $p < 0.001$ ) progressions after adjusting age at symptom onset as a confounding factor. It also reveals significant difference among four timepoints in both motor ( $F = 33.93$ ,  $p < 0.001$ ) and non-motor ( $F = 8.35$ ,  $p < 0.001$ ) progressions.

**Table 4** shows the comparison of the progression rates in three different time intervals between two subtypes. In terms of motor functioning, MMNS demonstrated more rapid early and long-term progression than the SMNS subtype. In terms of global cognitive function, measured using MoCA scores, patients in SMNS demonstrated rapid deterioration in early, secondary, and long-term progression than MMNS patients.



**FIGURE 2** | Distribution of cluster memberships in the motor-non-motor plane.



**FIGURE 3** | Level of differences in baseline characteristics between PD subtypes. The \* symbol indicates statistically significant differences between the two subtypes.

TABLE 2 | Brain imaging of PD subtypes.

Brain imaging	Mild motor and non-motor subtype (MMNS, N = 269) Mean (SD)	Severe motor and non-motor subtype (SMNS, N = 138) Mean (SD)	p-value ( $\alpha = 0.0017$ )
<b>Right caudate*</b>	2.083 ( $\pm 0.561$ )	1.841 ( $\pm 0.601$ )	<0.001
Left caudate	2.060 ( $\pm 0.563$ )	1.892 ( $\pm 0.625$ )	0.001
<b>Right putamen*</b>	0.882 ( $\pm 0.350$ )	0.771 ( $\pm 0.364$ )	<0.001
Left putamen	0.833 ( $\pm 0.345$ )	0.776 ( $\pm 0.375$ )	0.033
<b>Right striatum*</b>	2.966 ( $\pm 0.844$ )	2.612 ( $\pm 0.897$ )	<0.001
Left striatum	2.894 ( $\pm 0.841$ )	2.668 ( $\pm 0.954$ )	0.004
<b>Mean caudate measure*</b>	2.072 ( $\pm 0.514$ )	1.866 ( $\pm 0.574$ )	<0.001
<b>Mean putamen measure*</b>	0.858 ( $\pm 0.279$ )	0.773 ( $\pm 0.324$ )	<0.001
<b>Mean striatum measure*</b>	1.465 ( $\pm 0.369$ )	1.319 ( $\pm 0.425$ )	<0.001
<b>Lowest caudate measure*</b>	1.883 ( $\pm 0.517$ )	1.692 ( $\pm 0.539$ )	<0.001
<b>Highest caudate measure*</b>	2.261 ( $\pm 0.542$ )	2.040 ( $\pm 0.634$ )	<0.001
<b>Lowest putamen measure*</b>	0.691 ( $\pm 0.247$ )	0.638 ( $\pm 0.293$ )	<0.001
<b>Highest putamen measure*</b>	1.025 ( $\pm 0.355$ )	0.909 ( $\pm 0.387$ )	<0.001
<b>Lowest striatum measure*</b>	2.588 ( $\pm 0.699$ )	2.341 ( $\pm 0.783$ )	<0.001
<b>Highest striatum measure*</b>	3.271 ( $\pm 0.836$ )	2.938 ( $\pm 0.960$ )	<0.001
Asymmetry index (caudate)	19.047 ( $\pm 13.143$ )	18.822 ( $\pm 13.000$ )	0.854
Asymmetry index (putamen)	38.674 ( $\pm 25.199$ )	34.826 ( $\pm 24.438$ )	0.146
Asymmetry index striatum	23.549 ( $\pm 14.47$ )	22.644 ( $\pm 14.256$ )	0.529
Left count density ratio	2.707 ( $\pm 1.027$ )	2.645 ( $\pm 0.855$ )	0.854
Right count density ratio	2.547 ( $\pm 0.770$ )	2.619 ( $\pm 0.890$ )	0.638
<b>Contralateral striatum*</b>	2.622 ( $\pm 0.717$ )	2.376 ( $\pm 0.813$ )	<0.001
<b>Contralateral caudate*</b>	1.912 ( $\pm 0.530$ )	1.710 ( $\pm 0.561$ )	<0.001
Contralateral putamen	0.711 ( $\pm 0.254$ )	0.665 ( $\pm 0.300$ )	0.002
Contralateral count density ratio	2.885 ( $\pm 1.082$ )	2.758 ( $\pm 0.787$ )	0.348
<b>Ipsilateral striatum*</b>	3.237 ( $\pm 0.846$ )	2.904 ( $\pm 0.956$ )	<0.001
<b>Ipsilateral caudate*</b>	2.232 ( $\pm 0.547$ )	2.022 ( $\pm 0.624$ )	<0.001
<b>Ipsilateral putamen*</b>	1.005 ( $\pm 0.367$ )	0.882 ( $\pm 0.397$ )	<0.001
Ipsilateral count density ratio	2.367 ( $\pm 0.595$ )	2.507 ( $\pm 0.805$ )	0.285

\*Indicates statistical significance between the two subtypes (Wilcoxon rank-sum test).

## DISCUSSION

In the present study, we aimed to validate the Parkinson's disease subtypes using a standardized data-driven approach. We used k-means cluster analysis to unravel two distinct

TABLE 3 | CSF biomarkers of PD subtypes.

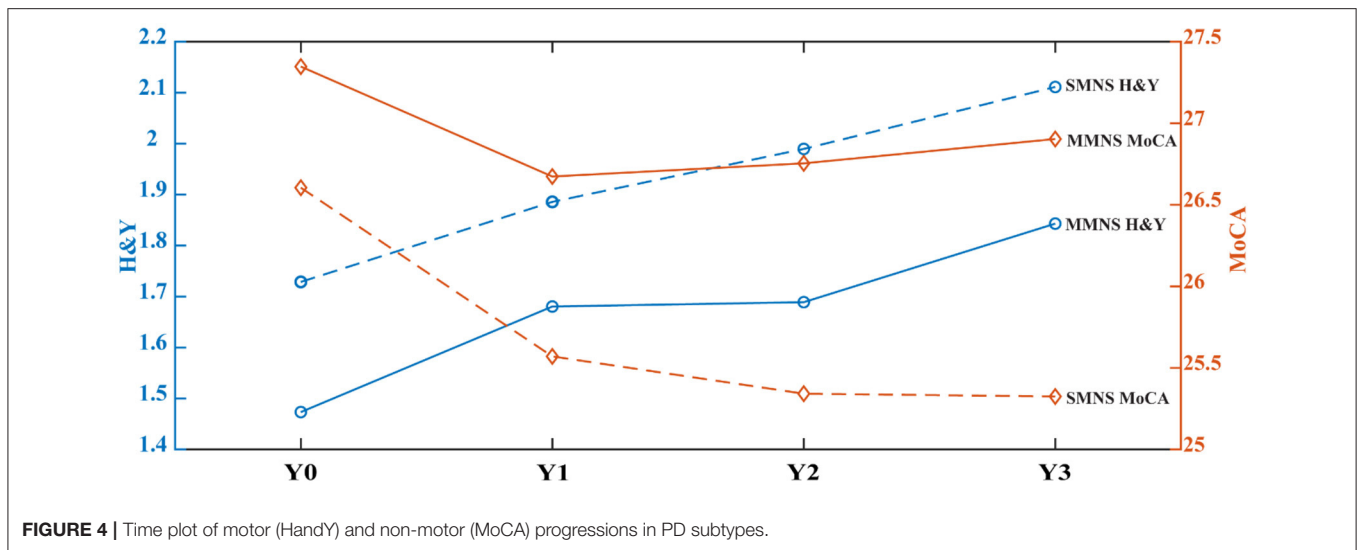
CSF variables	Mild motor and non-motor subtype (MMNS, N = 246–266) Mean (SD)	Severe motor and non-motor subtype (SMNS, N = 115–133) Mean (SD)	p-value ( $\alpha = 0.005$ )
A $\beta$ 1-42	920.070 ( $\pm 384.060$ )	908.880 ( $\pm 465.370$ )	0.529
$\alpha$ -synuclein	1,507.400 ( $\pm 652.890$ )	1,519.400 ( $\pm 713.130$ )	0.775
T-tau	166.100 ( $\pm 50.812$ )	177.710 ( $\pm 67.664$ )	0.260
P-tau	14.414 ( $\pm 4.649$ )	16.006 ( $\pm 6.296$ )	0.043
T-tau/A $\beta$ 1-42 ratio	0.1941 ( $\pm 0.0822$ )	0.219 ( $\pm 0.120$ )	0.080
P-tau/A $\beta$ 1-42 ratio	0.017 ( $\pm 0.008$ )	0.019 ( $\pm 0.012$ )	0.185
P-tau/T-tau ratio	0.085 ( $\pm 0.006$ )	0.085 ( $\pm 0.008$ )	0.624
A $\beta$ 1-42/ $\alpha$ -synuclein ratio	0.645 ( $\pm 0.214$ )	0.645 ( $\pm 0.307$ )	0.356
T-tau/ $\alpha$ -synuclein ratio	0.114 ( $\pm 0.026$ )	0.123 ( $\pm 0.038$ )	0.058
<b>P-tau/<math>\alpha</math>-synuclein ratio*</b>	0.009 ( $\pm 0.003$ )	0.010 ( $\pm 0.022$ )	0.002

\*Indicates statistically significant difference between two subtypes (Wilcoxon rank-sum test).

subtypes of Parkinson's disease patients using extensive motor and non-motor data from the Parkinson's Progression Markers Initiative (PPMI) database, which includes patients between 25 and 83 years, representing patients of broad age range and varying disease severity. We adopted a rigorous and standardized process to minimize biases in the cluster analysis, including data normalization, identifying optimal cluster numbers through an iterative process, and random initializations to avoid local optima. The process is fully data-driven without the need for manual parameter tuning such that objective and reproducible clustering results can be achieved. Based on the severity of motor and non-motor symptom burden, two distinct PD subtypes: severe motor-non-motor subtype (SMNS) and mild motor-non-motor subtype (MMNS) were identified. Novel motor and non-motor composite scores were developed to provide a direct and qualitative assessment of the distribution of cluster memberships. In addition, we provided a comprehensive evaluation of their characteristics, including brain imaging, CSF biomarkers, longitudinal motor and cognitive progression.

## Baseline Characteristics

Our *post-hoc* comparisons demonstrated that two subtypes varied significantly in terms of the age of symptoms onset, motor (MDS UPDRS III total score, rigidity, bradykinesia, and axial subscores) and non-motor (MoCA, SCOPA, RBD, GDS, pain, fatigue, and apathy) features. Although there is not much difference in the mean age of the two subtypes, the age range reflects the younger at symptom onset include patients of younger age (age range: 25.4–80.1 years) than older at symptom onset PD participants (age range: 35.6–83 years). These age ranges reflect the diversity in the age of *de novo* PD patients. The subtype with the older onset of symptoms experienced a greater



**TABLE 4 |** Parkinson's disease symptom progression rate of PD subtypes.

Characteristics	Mild motor and non-motor subtype (MMNS, N = 262) Mean (SD)	Severe motor and non-motor subtype (SMNS, N = 129) Mean (SD)
<b>Hoehn and Yahr score</b>		
- Early progression (Y1-Y0)	0.228 (±0.535)	0.171 (±0.528)
- Secondary progression (Y3-Y1)	0.169 (±0.587)	0.178 (±0.509)
- Long-term progression (Y3-Y0)	0.418 (±0.675)	0.383 (±0.699)
<b>MoCA score</b>		
- Early progression (Y1-Y0)	-0.693 (±2.573)	-1.031 (±2.783)
- Secondary progression (Y3-Y1)	0.254 (±2.104)	-0.274 (±2.571)
- Long-term progression (Y3-Y0)	-0.579 (±2.798)	-1.210 (±2.882)

Parkinson's symptoms burden than the subtype with the younger onset of symptoms. The SMNS subtype experienced more intense motor and non-motor symptoms than the MMNS subtype.

The severity of Parkinson's symptom in both groups were complemented by their respective SPECT imaging findings, which showed a marked reduction in binding values at caudate, putamen, and striatum regions in the SMNS relative to the MMNS subtype. These findings are in line with the earlier study findings, which demonstrated that reduced bindings values were indicative of Parkinson's disease (34). The intense Parkinson's symptom burden along with the shrinkage of caudate, putamen, striatum binding values in SMNS are suggestive of a pronounced dopaminergic neuronal loss in striatum nigral regions than MMNS groups (35). Earlier studies have indicated that greater loss of binding sites at caudate, putamen, and striatum regions were substantially correlated to impaired motor function, cognitive decline, and depression (21, 36, 37), which align with the current study findings.

Our findings from the CSF biomarkers comparison demonstrated that the two subtypes differ significantly regarding

the p-tau/ $\alpha$ -synuclein ratio. Our study is first to examine the correlations between the p-tau/ $\alpha$ -synuclein ratio with MoCA and UPDRS-III scores. The analyses showed that p-tau/ $\alpha$ -synuclein ratio was not significantly correlated with MoCA scores (Pearson correlation coefficient:  $\rho = 0.04$ ,  $p = 0.50$ ) or UPDRS-III (Pearson correlation coefficient:  $\rho = 0.06$ ,  $p = 0.23$ ). Despite insignificant findings, these findings suggest a critical need to examine the relationships between p-tau/ $\alpha$ -synuclein ratio with MoCA and UPDRS-III scores in the long run, as the progression of PD patients might illustrate some meaningful relationships. Prior studies have shown that a lower p-tau/ $\alpha$ -synuclein ratio is an important biomarker of PD and this ratio helps in differentiating PD from other neuropathies (30). However, the exact cut-off points of p-tau/ $\alpha$ -synuclein ratios for differentiating PD patients are unclear because of conflicting findings. Even though our findings are not directly comparable with previous findings, prior studies comparing p-tau/ $\alpha$ -synuclein ratios between PD patients and healthy cohorts show inconsistent findings. A higher p-tau/ $\alpha$ -synuclein ratios were observed in PD patients relative to healthy controls (38), whereas another study showed that p-tau/ $\alpha$ -synuclein ratios were not different between PD patients and healthy controls (39). It is important to note that the comparison groups of these studies vary from our PD subtypes, indicating the need to investigate the variation in p-tau/ $\alpha$ -synuclein ratios across PD subtypes. A study using PPMI data indicated that the p-tau/ $\alpha$ -synuclein ratios were not different across three identified subtypes of PD patients (17). Due to these discrepant findings, more studies are needed to investigate the role of the p-tau/ $\alpha$ -synuclein ratio and variation of the p-tau/ $\alpha$ -synuclein ratios across PD subtypes.

## Rate of Progression

Our findings on the rate of progression show that MMNS showed more rapid early and long-term deterioration of motor decline than SMNS as measured by Hoehn and Yahr score. In terms of global cognitive functioning, the SMNS group deteriorated more rapidly than MMNS patients did. Our findings confirmed the



degenerative, progressive nature of Parkinson's disease in terms of both motor and cognitive domains. As depicted in **Figure 4**, clear contrast can be drawn in the progression of motor and non-motor symptoms between two subtypes. These findings shed light on unique trajectories of motor and cognitive dysfunctions in these two subtypes. Moreover, the ranking of severity in identified subtypes presents a stable pattern for both motor- and non-motor symptoms up to 3-year following the initial baseline visit, indicating the temporally consistent property of these subtypes, and demonstrating their potential prognostic value.

## Comparison With Previous Studies

Our findings are consistent with a study by Pagano et al. (40) using PPMI data with a priori age-based clusters. This study demonstrated that Parkinson's disease patients (older at symptom onset) manifested severe motor symptoms (tremor, MDS-UPDRS III total score, and bradykinesia), non-motor symptoms (SCOPA total, UPSIT, and MoCA), and marked deterioration in imaging makers than the younger at symptom onset (40). Our findings demonstrating the relationship between age of symptom onset and disease severity also agrees with previous studies by Fereshtehnejad et al. (17) and Erro et al. (18) based on data-driven analyses—without a priori hypotheses, using PPMI data. Especially, the SMNS subtype—that experienced more intense Parkinson's symptoms—bears some resemblance with their diffuse malignant subtype in respect of motor (MDS-UPDRS Part III score), non-motor (SCOPA-total, MoCA), and neuropsychiatric features (apathy, fatigue, pain, STAI). Our MMNS subtype with milder symptoms shares similarities with benign groups in terms of motor (MDS-UPDRS Part III score, tremor), non-motor (SCOPA-total, MoCA), and neuropsychiatric features (STAI, hallucinations, apathy, pain). Although the above-mentioned studies obtained their data from the common source, PPMI, findings reveal some discrepancies with respect to the motor and non-motor symptoms that might be due to the use of different clustering analysis techniques.

The past studies have reported that tremor is one of the key distinguishing features across various subtypes, whereas the current study showed that two subtypes did not vary in terms of tremor severity. Thus, our findings do not align with some past studies that identified unique subtypes with tremor and non-tremor dominant features (41, 42). There are several plausible reasons explaining differences in our findings with these studies. One reason could be that most of these studies utilized motor data alone, while our study considered both motor and non-motor symptoms data of the PD. The second plausible reason might be due to consideration of cross-sectional tremor data in the cluster analysis that might have prevented tracking of tremor progression and discrepancies in identified subtypes. Finally, it might be because of PPMI data as PPMI only includes *de novo* PD patients who are off the medications, which might have excluded the PD who were on medications with greater tremor severity. The disease progression rates in this current study were also inconsistent with prior studies, which report that the old-age onset subtype showed rapid motor decline than the young-onset subtype having slower progression of motor decline (41, 42). The likely reasons for these inconsistencies can

be due to differences in variables included in cluster analyses, data pre-processing, data standardization, and clustering techniques applied for identifying clusters.

## Strengths and Limitations

The present study developed a standardized cluster analysis procedure that builds upon recommendations and common practice from previous studies (4, 6, 12, 18), and it is published online to foster data comparison and reproducibility of results. We undertook measures to mitigate the inherent limitations associated with data-driven approach analysis: (a) inclusion of comprehensive motor and non-motor features from the publicly available PPMI database, (b) data standardization using min-max normalization, (c) identification and selection of optimum clusters using C-H pseudo-F value, (d) identification of global minima and avoidance of local minima by setting iterations and repetitions to 100 levels. These measures aid in minimizing any subjective biases that might have stemmed from a priori hypothesis-driven analysis, clinical subtyping, and the heuristic selection of variables and clusters. Our study explored and compared attributes of two clusters using motor and non-motor features, brain imaging, CSF-biomarkers, trajectories of motor, and global cognitive functioning. The novel composite scores combined with visualization of cluster memberships in the two-dimensional motor-non-motor plane provide a more intuitive and clinically meaningful understanding of identified PD subtypes. In addition, few studies have investigated the longitudinal stability of subtype characteristics. Our results show a clear and stable contrast between the two subtypes in both motor and cognitive progressions. Such stability provides not only additional validation but also offers great clinical implications in terms of the prognostic value of the identified PD subtypes.

The last strength of the study inherits from the advantage of the data-driven cluster analysis in that once clusters have been identified, a rule for assigning cluster memberships (i.e., PD subtypes in this study) is readily constructed and can be used to classify patient data into one of these two PD subtypes. Our K-means cluster analysis approach derives rescaling parameters for normalizing patient data and centroid information for the two PD subtypes (see **Supplementary Table S1**). These parameters collectively form the cluster model, which can classify patients into one of two PD subtypes based on the following steps, (a) normalize patient data using rescaling parameters; (b) calculate squared Euclidean distances between patient data and each cluster centroid; (c) classify the patient into the PD subtype with a shorter distance. The simple implementation and explainable model parameters facilitate the downstream clinical applications of identified PD subtypes. Compared to traditional clinical subtyping that is usually based on a priori hypothesis and relies on motor symptoms and/or the age of onset (10, 11, 43–46), data-driven PD subtyping addresses some of the limitations. This includes incorporating important non-motor features in the subtyping process and bypassing the ambiguity in selecting the ratio of motor sub-scores for defining the clinical subtypes (12, 47). However, several challenges still exist, staggering the broad adoption of data-driven subtypes. First, data-driven PD

subtyping often requires more data elements to be collected, which imposes an additional burden in clinical practice. Second, there is still a substantial variation between subtypes identified from different data-driven PD subtyping studies. Our study answers the call for a standardized cluster analysis approach (12) and facilitates result sharing and comparison. More efforts need to be directed in this direction to identify the most relevant variables and consolidate discrepant findings for data-driven PD subtyping to push their clinical use.

Despite the strengths in the study, our findings should be interpreted with some degree of caution partially due to limitations associated with PPMI datasets. PPMI involves untreated PD patients at the early PD stage without extreme motor or cognitive complications. This might underrepresent patients with greater symptoms severity who are on medications. PPMI has a shorter follow-up duration (<4 years); thus, it was not possible to examine the long trajectories of disease severity across subtypes. Furthermore, confounding factors in longitudinal analyses in observational studies often exist. Although the issue is mitigated by including the age at symptom onset as a covariate in the ANCOVA test, other potential confounding factors have yet to be explored due to the limitations of existing datasets. For example, procedures such as deep brain stimulation (DBS) are likely to alter the symptom progression trajectories. However, among 391 patient data entered the longitudinal analysis, only 48 of them have information on whether and when DBS procedures are administrated. Therefore, future study is warranted to investigate the confounding effects further when more complete datasets become available.

Although PPMI includes patients of a wider age range and disease severity, we chose to exclude the participants with any missing values in variables used for cluster analysis to prevent potential biases that might be introduced due to the replacing of missing values. Due to greater missing values in imaging and CSF biomarkers, several participants were excluded from analyses that might have compromised the generalizability of findings. This exclusion of participants with missing values might have restricted the representativeness of the PPMI dataset. The inclusion of participants with complete data for cluster analysis might have reduced the strength of the data-driven analysis due to the truncation of the analytic sample size. In addition, we only summarize clinical, brain imaging, and CSF biomarker characteristics of two subtypes using descriptive measures; thus, this might not explain the correlation between clinical, imaging, and biomarker changes.

As the PPMI dataset is still evolving with the inclusion of motor and non-motor features and longer follow-up duration, further confirmatory studies are required to validate PD subtypes. Identification of subtypes of PD patients using the latest, complete PPMI data will help in the precise identification of PD subtypes representing PD patients. Furthermore, subtypes can be identified using brain imaging, biomarkers, and genetics data. Moreover, future studies need to examine the correlation between clinical features, CSF, and imaging markers to elucidate the complex heterogeneous nature of Parkinson's disease. Other

clustering approaches, such as network analysis, can be used to validate PD subtypes.

## CONCLUSION

In summary, we identified two distinct subtypes of Parkinson's disease: (a) mild motor-non-motor subtype (MMNS) with milder motor, non-motor, and neuropsychiatric features and (b) severe motor-non-motor subtype (SMNS) with intense motor and non-motor features by including motor and non-motor features in cluster analysis. The SPECT imaging supports the clinical symptom burden—lower binding values in striatal pathways were observed in SMNS than MMNS, reflecting greater deterioration in nigral pathways in the SMNS subtype. Both subtypes showed peculiar progression patterns, with SMNS showing a more rapid decline in global cognitive functioning and MMNS showing a more rapid decline in motor functioning. The clear contrast between two subtypes in both motor and cognitive functioning progressions demonstrates the longitudinal stability of subtype memberships up to 3 years following the baseline visit. Such stability provides great clinical implications on the prognostic value of the identified subtypes at baseline. More studies are needed to confirm our findings with a larger sample of longitudinal data for subtyping analysis. Such findings provide additional evidence to devise tailored treatment planning for identified PD subtypes to maximize patient outcomes.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study was obtained from the Parkinson's Progressive Marker Initiative (PPMI; <https://www.ppmi-info.org/>), the following licenses/restrictions apply: Investigators seeking access to PPMI data must submit an Online Application, sign the Data Use Agreement and comply with the study Publications Policy. Requests to access these datasets should be directed to the PPMI Data and Publications Committee (DPC), <https://www.ppmi-info.org/access-data-specimens/download-data>.

## AUTHOR CONTRIBUTIONS

RX and XH contributed to the conception, design of the work, revised the work, and have approved the submitted version. SS, JP, and RX contributed to the analysis and interpretation of the data. SS drafted the work. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We graciously thank Dr. Qing Yang for her suggestions on statistical tests in this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.810038/full#supplementary-material>

## REFERENCES

- Thenganatt MA, Jankovic J. Parkinson disease subtypes. *JAMA Neurol.* (2014) 71:499–504. doi: 10.1001/jamaneurol.2013.6233
- Marsili L, Rizzo G, Colosimo C. Diagnostic criteria for Parkinson's disease: from James Parkinson to the concept of prodromal disease. *Front Neurol.* (2018) 9:156. doi: 10.3389/fneur.2018.00156
- Sauerbier A, Jenner P, Todorova A, Chaudhuri KR. Non motor subtypes and Parkinson's disease. *Parkinsonism Relat Disord.* (2016) 22:S41–46. doi: 10.1016/j.parkreldis.2015.09.027
- Fereshtehnejad SM, Postuma RB. Subtypes of Parkinson's disease: what do they tell us about disease progression? *Curr Neurol Neurosci Rep.* (2017) 17:34. doi: 10.1007/s11910-017-0738-x
- Mu J, Chaudhuri KR, Bielza C, de Pedro-Cuesta J, Larrañaga P, Martínez-Martin P. Parkinson's disease subtypes identified from cluster analysis of motor and non-motor symptoms. *Front Aging Neurosci.* (2017) 9:301. doi: 10.3389/fnagi.2017.00301
- Van Rooden S, Heiser WJ, Kok JN, Verbaan D, Van Hilten JJ, Marinus J. The identification of Parkinson's disease subtypes using cluster analysis: a systematic review. *Mov Disord.* (2010) 25:969–78. doi: 10.1002/mds.23116
- Jankovic J, Tan EK. Parkinson's disease: etiopathogenesis and treatment. *J Neurol Neurosurg Psychiatry.* (2020) 91:795–808. doi: 10.1136/jnnp-2019-322338
- Marras C, Chaudhuri KR, Titova N, Mestre TA. Therapy of Parkinson's disease subtypes. *Neurotherapeutics.* (2020) 17:1366–77. doi: 10.1007/s13311-020-00894-7
- Marras C, Chaudhuri KR. Nonmotor features of Parkinson's disease subtypes. *Mov Disord.* (2016) 31:1095–102. doi: 10.1002/mds.26510
- Jankovic J, McDermott M, Carter J, Gauthier S, Goetz C, Golbe L, et al. Variable expression of Parkinson's disease: a baseline analysis of the DATATOP cohort. The Parkinson Study Group. *Neurology.* (1990) 40:1529–34. doi: 10.1212/WNL.40.10.1529
- Zettsky WJ, Jankovic J, Pirozzolo FJ. The heterogeneity of Parkinson's disease: clinical and prognostic implications. *Neurology.* (1985) 35:522–6. doi: 10.1212/WNL.35.4.522
- Qian E, Huang Y. Subtyping of Parkinson's disease - where are we up to? *Aging Dis.* (2019) 10:1130–9. doi: 10.14336/AD.2019.0112
- Graham JM, Sagar HJ. A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes. *Mov Disord.* (1999) 14:10–20. doi: 10.1002/1531-8257(199901)14:1<10::aid-mds1005>3.0.co;2-4
- Erro R, Vitale C, Amboni M, Picillo M, Moccia M, Longo K, et al. The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients. *PLoS ONE.* (2013) 8:e70244. doi: 10.1371/journal.pone.0070244
- Fereshtehnejad SM, Romenets SR, Anang JB, Latreille V, Gagnon JF, Postuma RB. New clinical subtypes of parkinson disease and their longitudinal progression: a prospective cohort comparison with other Phenotypes. *JAMA Neurol.* (2015) 72:863–73. doi: 10.1001/jamaneurol.2015.0703
- Van Rooden SM, Colas F, Martínez-Martin P, Visser M, Verbaan D, Marinus J, et al. Clinical subtypes of Parkinson's disease. *Mov Disord.* (2011) 26:51–8. doi: 10.1002/mds.23346
- Fereshtehnejad SM, Zeighami Y, Dagher A, Postuma RB. Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. *Brain.* (2017) 140:1959–76. doi: 10.1093/brain/awx118
- Erro R, Picillo M, Vitale C, Palladino R, Amboni M, Moccia M, et al. Clinical clusters and dopaminergic dysfunction in *de-novo* Parkinson disease. *Parkinsonism Relat Disord.* (2016) 28:137–40. doi: 10.1016/j.parkreldis.2016.04.026
- Mestre TA, Eberly S, Tanner C, Grimes D, Lang AE, Oakes D, et al. Reproducibility of data-driven Parkinson's disease subtypes for clinical research. *Parkinsonism Relat Disord.* (2018) 56:102–6. doi: 10.1016/j.parkreldis.2018.07.009
- Marek K, Chowdhury S, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, et al. The Parkinson's progression markers initiative (PPMI)—establishing a PD biomarker cohort. *Ann Clin Transl Neurol.* (2018) 5:460–77. doi: 10.1002/acn3.644
- Bayram E, Kaplan N, Shan G, Caldwell JZ. The longitudinal associations between cognition, mood and striatal dopaminergic binding in Parkinson's Disease. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn.* (2020) 27:581–94. doi: 10.1080/13825585.2019.1653445
- Li X, Xing Y, Martin-Bastida A, Piccini P, Auer DP. Patterns of grey matter loss associated with motor subscores in early Parkinson's disease. *NeuroImage Clin.* (2018) 17:498–504. doi: 10.1016/j.nicl.2017.11.009
- Antonini A, Benti R, De Notaris R, Tesi S, Zecchinelli A, Sacilotto G, et al. <sup>123</sup>I-Ioflupane/SPECT binding to striatal dopamine transporter (DAT) uptake in patients with Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy. *Neurol Sci.* (2003) 24:149–50. doi: 10.1007/s10072-003-0103-5
- Benamer HT, Patterson J, Grosset DG, Booij J, De Bruin K, Van Royen E, et al. Accurate differentiation of parkinsonism and essential tremor using visual assessment of [<sup>123</sup>I]-FP-CIT SPECT imaging: the [<sup>123</sup>I]-FP-CIT study group. *Mov Disord.* (2000) 15:503–10. doi: 10.1002/1531-8257(200005)15:3<503::AID-MDS1013>3.0.CO;2-V
- Leaver K, Poston KL. Do CSF biomarkers predict progression to cognitive impairment in Parkinson's disease patients? A systematic review. *Neuropsychol Rev.* (2015) 25:411–23. doi: 10.1007/s11065-015-9307-8
- Zhang Y, Wu I-W, Tosun D, Foster E, Schuff N. Progression of regional microstructural degeneration in Parkinson's disease: a multicenter diffusion tensor imaging study. *PLoS ONE.* (2016) 11:e0165540. doi: 10.1371/journal.pone.0165540
- Katayama T, Sawada J, Takahashi K, Yahara O. Cerebrospinal fluid biomarkers in Parkinson's disease: a critical overview of the literature and meta-analyses. *Brain Sci.* (2020) 10:466. doi: 10.3390/brainsci10070466
- Mollenhauer B, Caspell-Garcia CJ, Coffey CS, Taylor P, Singleton A, Shaw LM, et al. Longitudinal analyses of cerebrospinal fluid alpha-Synuclein in prodromal and early Parkinson's disease. *Mov Disord.* (2019) 34:1354–64. doi: 10.1002/mds.27806
- Majbour NK, Vaikath NN, Eusebi P, Chiasserini D, Ardah M, Varghese S, et al. Longitudinal changes in CSF alpha-synuclein species reflect Parkinson's disease progression. *Mov Disord.* (2016) 31:1535–42. doi: 10.1002/mds.26754
- Parnetti L, Chiasserini D, Bellomo G, Giannandrea D, De Carlo C, Qureshi MM, et al. Cerebrospinal fluid Tau/alpha-synuclein ratio in Parkinson's disease and degenerative dementias. *Mov Disord.* (2011) 26:1428–35. doi: 10.1002/mds.23670
- Martinez-Martin P, Skorvanek M, Rojo-Abuin JM, Gregova Z, Stebbins GT, Goetz CG, et al. Validation study of the hoehn and yahr scale included in the MDS-UPDRS. *Mov Disord.* (2018) 33:651–2. doi: 10.1002/mds.27242
- Zhao YJ, Wee HL, Chan YH, Seah SH, Au WL, Lau PN, et al. Progression of Parkinson's disease as evaluated by Hoehn and Yahr stage transition times. *Mov Disord.* (2010) 25:710–6. doi: 10.1002/mds.22875
- Hentschke H. Available online at: <https://github.com/hhentschke/measures-of-effect-size-toolbox>, GitHub (accessed December 20, 2021).
- Kaasinen V, Vahlberg T, Stoessl AJ, Strafella AP, Antonini A. Dopamine receptors in Parkinson's disease: a meta-analysis of imaging studies. *Mov Disord.* (2021) 36:1781–91. doi: 10.1002/mds.28632
- Dauer W, Przedborski S. Parkinson's disease: mechanisms and models. *Neuron.* (2003) 39:889–909. doi: 10.1016/S0896-6273(03)00568-3
- Pasquini J, Durcan R, Wiblin L, Stokholm MG, Rochester L, Brooks DJ, et al. Clinical implications of early caudate dysfunction in Parkinson's disease. *J Neurol Neurosurg Psychiatry.* (2019) 90:1098–104. doi: 10.1136/jnnp-2018-320157
- Benamer HT, Patterson J, Wyper DJ, Hadley DM, Macphee G, Grosset DG. Correlation of Parkinson's disease severity and duration with 123I-FP-CIT SPECT striatal uptake. *Mov Disord.* (2000) 15:692–8. doi: 10.1002/1531-8257(200007)15:4<692::aid-mds1014>3.0.co;2-v
- Delgado-Alvarado M, Gago B, Gorostidi A, Jiménez-Urbietta H, Dacosta-Aguayo R, Navalpotro-Gómez I, et al. Tau/alpha-synuclein ratio and inflammatory proteins in Parkinson's disease: an exploratory study. *Mov Disord.* (2017) 32:1066–73. doi: 10.1002/mds.27001
- Kang JH, Mollenhauer B, Coffey CS, Toledo JB, Weintraub D, Galasko DR, et al. CSF biomarkers associated with disease heterogeneity in early Parkinson's disease: the Parkinson's progression markers initiative study. *Acta Neuropathol.* (2016) 131:935–49. doi: 10.1007/s00401-016-1552-2

40. Pagano G, Ferrara N, Brooks DJ, Pavese N. Age at onset and Parkinson disease phenotype. *Neurology*. (2016) 86:1400–7. doi: 10.1212/WNL.0000000000002461
41. Lewis S, Foltynie T, Blackwell AD, Robbins TW, Owen AM, Barker RA. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. *J Neurol Neurosurg Psychiatry*. (2005) 76:343–8. doi: 10.1136/jnnp.2003.033530
42. Reijnders JS, Ehrt U, Lousberg R, Aarsland D, Leentjens AF. The association between motor subtypes and psychopathology in Parkinson's disease. *Parkinsonism Relat Disord*. (2009) 15:379–82. doi: 10.1016/j.parkreldis.2008.09.003
43. Salawu FK. Patient considerations in early management of Parkinson's disease: focus on extended-release pramipexole. *Patient Prefer Adherence*. (2012) 6:49–54. doi: 10.2147/PPA.S11841
44. Alves G, Wentzel-Larsen T, Aarsland D, Larsen JP. Progression of motor impairment and disability in Parkinson disease: a population-based study. *Neurology*. (2005) 65:1436–41. doi: 10.1212/01.wnl.0000183359.50822.f2
45. Ransmayr G, Poewe W, Ploerer S, Birbamer G, Gerstenbrand F. Psychometric findings in clinical subtypes of Parkinson's disease. *Adv Neurol*. (1987) 45:409–411.
46. Schiess MC, Zheng H, Soukup VM, Bonnen JG, Nauta HJ. Parkinson's disease subtypes: clinical classification and ventricular cerebrospinal fluid analysis. *Parkinsonism Relat Disord*. (2000) 6:69–76. doi: 10.1016/S1353-8020(99)00051-6
47. von Coelln R, Shulman LM. Clinical subtypes and genetic heterogeneity: of lumping and splitting in Parkinson disease. *Curr Opin Neurol*. (2016) 29:727–34. doi: 10.1097/WCO.0000000000000384

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Shakya, Prevett, Hu and Xiao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.