doi:10.1093/brain/awx118 BRAIN 2017: 140; 1959–1976 | 1959

BRAIN
A JOURNAL OF NEUROLOGY

# Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression

**Seyed-Mohammad Fereshtehnejad,[1,2] Yashar Zeighami,[3] Alain Dagher[3] and Ronald B. Postuma[1,4]**

Parkinson's disease varies widely in clinical manifestations, course of progression and biomarker profiles from person to person. Identification of distinct Parkinson's disease subtypes is of great priority to illuminate underlying pathophysiology, predict progression and develop more efficient personalized care approaches. There is currently no clear way to define and divide subtypes in Parkinson's disease. Using data from the Parkinson's Progression Markers Initiative, we aimed to identify distinct subgroups via cluster analysis of a comprehensive dataset at baseline (i.e. cross-sectionally) consisting of clinical characteristics, neuroimaging, biospecimen and genetic information, then to develop criteria to assign patients to a Parkinson's disease subtype. Four hundred and twenty-one individuals with *de novo* early Parkinson's disease were included from this prospective longitudinal multicentre cohort. Hierarchical cluster analysis was performed using data on demographic and genetic information, motor symptoms and signs, neuropsychological testing and other non-motor manifestations. The key classifiers in cluster analysis were a motor summary score and three non-motor features (cognitive impairment, rapid eye movement sleep behaviour disorder and dysautonomia). We then defined three distinct subtypes of Parkinson's disease patients: 223 patients were classified as 'mild motor-predominant' (defined as composite motor and all three non-motor scores below the 75th percentile), 52 as 'diffuse malignant' (composite motor score plus either $\geq$1/3 non-motor score $>$75th percentile, or all three non-motor scores $>$75th percentile) and 146 as 'intermediate'. On biomarkers, people with diffuse malignant Parkinson's disease had the lowest level of cerebrospinal fluid amyloid-β (329.0 $\pm$ 96.7 pg/ml, $P = 0.006$) and amyloid-β/total-tau ratio (8.2 $\pm$ 3.0, $P = 0.032$). Data from deformation-based magnetic resonance imaging morphometry demonstrated a Parkinson's disease-specific brain network had more atrophy in the diffuse malignant subtype, with the mild motor-predominant subtype having the least atrophy. Although disease duration at initial visit and follow-up time were similar between subtypes, patients with diffuse malignant Parkinson's disease progressed faster in overall prognosis (global composite outcome), with greater decline in cognition and in dopamine functional neuroimaging after an average of 2.7 years. In conclusion, we introduce new clinical criteria for subtyping Parkinson's disease based on a comprehensive list of clinical manifestations and biomarkers. This clinical subtyping can now be applied to individual patients for use in clinical practice using baseline clinical information. Even though all participants had a recent diagnosis of Parkinson's disease, patients with the diffuse malignant subtype already demonstrated a more profound dopaminergic deficit, increased atrophy in Parkinson's disease brain networks, a more Alzheimer's disease-like cerebrospinal fluid profile and faster progression of motor and cognitive deficits.

1 Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada
2 Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society (NVS), Karolinska Institutet, Stockholm, Sweden
3 McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada
4 Centre for Advanced Research in Sleep Medicine, Hôpital du Sacré-Coeur de Montréal, Montreal, QC, Canada

Received December 16, 2016. Revised March 2, 2017. Accepted March 29, 2017. Advance Access publication May 25, 2017

Correspondence to: Dr Ronald B. Postuma
Department of Neurology, L7-305 Montreal General Hospital, 1650 Cedar Avenue, Montreal, Quebec, Canada H3G 1A4
E-mail: ron.postuma@mcgill.ca

**Keywords:** Parkinson's disease; subtypes; biomarkers; neuroimaging; prognosis

**Abbreviations:** GCO = global composite outcome; MDS-UPDRS = Movement Disorder Society-Unified Parkinson's Disease Rating Scale; MoCA = Montreal Cognitive Assessment; PIGD = postural instability–gait difficulty; PPMI = Parkinson's Progression Markers Initiative; RBD = REM sleep behaviour disorder

# Introduction

Parkinson's disease varies considerably in its clinical manifestations and prognosis, suggesting it may be divisible into subtypes (Berg *et al.*, 2014). Recently the National Institutes of Health established subtype identification as one of the top three clinical research priorities in Parkinson's disease (Sieber *et al.*, 2014). Defining different subcategories of Parkinson's disease is key to better understand underlying disease mechanisms, predict disease course, and design clinical trials. Yet, the means to identify subtypes and predict individual prognosis remain undefined.

Disease subtypes can be identified with cluster analyses, which use a hypothesis-free data-driven approach. Previous studies used cluster analysis to define clinical Parkinson's disease subtypes based on motor severity, motor complications, certain non-motor features and demographic characteristics (van Rooden *et al.*, 2010; Fereshtehnejad *et al.*, 2015). Depth of phenotypic information in these studies was variable, and often limited. Moreover, all previous cluster analyses were only based on clinical features; neither neuroimaging, nor serum/CSF biomarkers nor genetic data were available. Many previous studies were limited by insufficient longitudinal assessments to evaluate prognosis of subtypes. Finally, the final output of statistical approaches such as clustering can be described at the group level only. In order to be used in practice, statistical subtyping solutions need to be translated into a method to assign individual patients to a subtype.

The Parkinson's Progression Markers Initiative (PPMI) is a comprehensive longitudinal, international multi-centre database consisting of clinical, genetic, neuroimaging, and blood/CSF biomarkers of over 400 *de novo* Parkinson's disease patients (PPMI, 2011). All clinical features are also annually reassessed. This extensive phenotypic and biomarker information provides a unique opportunity to assess the heterogeneity of Parkinson's disease, to create the most comprehensive definition of subtypes yet performed, and to allow longitudinal assessment of disease progression and outcome of different subtypes.

The aims of our study were to: (i) perform cluster analysis on a comprehensive baseline dataset, including both motor and non-motor clinical characteristics and genetic information to identify distinct Parkinson's disease clusters; (ii) introduce a new practical classification method to assign individual patients to their subtype; (iii) assess on *post hoc* analyses the neuroimaging, biospecimen and clinical characteristics of each Parkinson's disease subtype; and (iv) compare disease progression between different Parkinson's disease subtypes.

# Materials and methods

## Study setting and population

PPMI (http://www.ppmi-info.org) has been extensively described elsewhere (PPMI, 2011). Recruitment criteria include: age $\geqslant 30$, Parkinson's disease diagnosis within the last 2 years, baseline Hoehn and Yahr Stage I–II, and no anticipated need for symptomatic treatment within 6 months of baseline (PPMI, 2011). The institutional review board approved the PPMI protocol in all participating sites. Written informed consent was attained from all participants.

We obtained data from the PPMI database on May 2016 in compliance with the PPMI Data Use Agreement. Any individual with >20% missing values on baseline data was excluded ($n = 421$ included).

## Baseline and clinical assessments

In PPMI, a comprehensive set of clinical features including both motor and non-motor symptoms is assessed. We used the following data:

(i) Demographics: age, sex, race, family history, symptom duration, education level.
(ii) Blood biomarkers: biochemical tests.
(iii) Motor manifestations: International Parkinson's disease and Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS)-Part II, MDS-UPDRS-Part III, total tremor score, postural instability–gait difficulty (PIGD) score, tremor/PIGD motor phenotype (Stebbins *et al.*, 2013), Schwab-England activities of daily living (ADL) score.
(iv) Neuropsychological features: age/education adjusted Montreal Cognitive Assessment (MoCA) (Nasreddine *et al.*, 2005), all neuropsychological variables including visuospatial function [Benton Judgment of Line Orientation (JOLO)] (Benton *et al.*, 1978), speed/attention (Symbol–Digit Matching) (Smith, 1991), memory [Hopkins Verbal Learning Test (HVLT) for total recall, delayed recall, retention and recognition-discrimination] (Shapiro *et al.*, 1999) and executive function [semantic verbal-language fluency test (Gladsjo *et al.*, 1999) and letter-number sequencing (LNS) (Wechsler, 2008)]. Baseline analysis of

cognitive performance in PPMI has been recently reported (Weintraub *et al.*, 2015).

(v) Other non-motor manifestations: MDS-UPDRS-Part I, olfactory dysfunction [age/sex adjusted University of Pennsylvania Smell Identification Test (UPSIT) score] (Doty *et al.*, 1984), autonomic dysfunction [Scales for Outcomes in Parkinson's disease-Autonomic (SCOPA-AUT) total score and its subsections: orofacial, constipation, urinary, cardiovascular, thermoregulatory, pupillo-motor and sexual] (Visser *et al.*, 2004), orthostatic drop in systolic blood pressure, depression [Geriatric Depression Scale (GDS) score] (Yesavage and Sheikh, 1986), anxiety [State-Trait Anxiety Inventory (STAI) score] (Spielberger *et al.*, 1983), REM sleep behaviour disorder (RBD) [RBD screening questionnaire (RBDSQ) score] (Stiasny-Kolster *et al.*, 2007), sleep disturbances [Epworth Sleepiness Score (ESS)] (Johns, 1991), impulse control disorders (ICD) [Questionnaire for Impulsive-Compulsive Disorders in Parkinson's disease (QUIP) score] (Weintraub *et al.*, 2009). For other non-motor features without specific tools for measurement (hallucinations, apathy, pain, fatigue) we used single items from MDS-UPDRS-Part I.

## Genetic information

A single summary indicator genetic risk score was used. This was previously calculated in PPMI by summing the number of 28 common risk loci identified in a large-scale meta-analysis of Parkinson's disease genome-wide association studies, plus two additional risk variants detected within PPMI (p.N370S in GBA and p.G2019S in LRRK2) (Nalls *et al.*, 2015).

## CSF biomarkers

CSF amyloid-$\beta_{1\text{-}42}$, total (t)-tau, and phosphorylated tau (P-tau181) were measured by INNO-BIA AlzBio3 immunoassay (Innogenetics Inc.), and α-synuclein was measured by enzyme-linked immunosorbent assay (Kang *et al.*, 2013). We also calculated ratios between CSF biomarkers as recently recommended (Kang *et al.*, 2016).

## Imaging biomarkers

Single-photon emission computed tomography (SPECT) with the DAT tracer $^{123}$I-ioflupane was obtained in 351 PPMI participants at baseline and follow-up visits (PPMI, 2011). The striatal binding ratio using the occipital lobe as a reference region, was calculated for the left and right caudate and putamen separately. High resolution $T_1$-weighted 3 T MRI were available for 229 individuals with Parkinson's disease and 117 healthy age-matched controls at baseline. Deformation-based morphometry (DBM) was used as a measure of brain atrophy. The analysis was performed as described previously (Zeighami *et al.*, 2015). Briefly, each subject's MRI was first linearly and then non-linearly registered to the Montreal Neurological Institute (MNI) ICBM-152 template. Non-linear transformations were used to calculate the Jacobian determinant of the deformation matrix at each voxel for each subject. We then performed independent component analysis (ICA) on the DBM maps to identify Parkinson's disease-specific atrophy networks in early Parkinson's disease (Zeighami *et al.*, 2015). Two indicators of atrophy were calculated for each participant: the Parkinson's disease-related network atrophy score from the DBM ICA, and the substantia nigra atrophy score, equal to the DBM measure in the substantia nigra. Because MRI was only performed in 229 patients (and also to enhance clinical applicability), MRI variables were not included in the clustering solution, but were tested in *post hoc* comparisons.

## Longitudinal assessments

As all the above clinical and neuroimaging features are re-assessed annually (PPMI, 2011), we also analysed progression over time. Follow-up time is variable; we selected the earliest and latest recorded data for each feature, with the interval between the two measurements calculated for each individual. Patients with less than 1-year follow-up were excluded from longitudinal analysis, leaving 401 cases analysed for progression.

## Global composite outcome

Parkinson's disease has diverse manifestations, all of which contribute to quality of life and overall disease progression. For analysis of progression, we created a global composite outcome (GCO) as a single numeric indicator of prognosis, similar to our previous single-centre study on Parkinson's disease clustering (Fereshtehnejad *et al.*, 2015). This merged the most important clinical domains into a single aggregate, simultaneously accounting for different variations in the range/direction of scores, while avoiding overweighting a single domain. The GCO equally weighted non-motor symptoms (UPDRS-I), motor symptoms (UPDRS-II), motor signs (UPDRS-III), overall activities of daily living (Schwab and England ADL), and global cognition (MoCA), standardized by averaging the z-scores of each component. For calculating change we used the mean/SD from baseline as reference, to assess progression. Calculations for each component were:

$$\text{Baseline}: \text{z-score}_{\text{baseline}} = (\text{crude score}_{\text{baseline}} - \text{mean}_{\text{baseline}})/\text{SD}_{\text{baseline}}$$

$$\text{Follow-up}: \text{z-score}_{\text{follow-up}} = (\text{crude score}_{\text{follow-up}} - \text{mean}_{\text{baseline}})/\text{SD}_{\text{baseline}}$$

The total GCO was calculated by averaging all components' z-scores (higher GCO scores indicate worse function).

## Statistical methods and cluster/subtype definition

### Data preparation and missing data imputation

We used *R* version 3.2.2 (https://www.r-project.org) to read variables and extract scores from the PPMI database (Supplementary material). Where appropriate, scores were calculated based on normative values (for instance, using %normal age/sex adjusted UPSIT rather than crude scores). Missing values (<4%) were imputed by using mean values for the entire cohort.

### Cluster analysis

To avoid overweighting a single feature or single domain in the clustering solution, we generated composite indicators, each of which summarizes several redundant variables

(for instance the composite cognitive memory score used averaged z-scores for all single HVLT scores). This also provided dimension reduction, allowing the number of statistical variables to better correspond with the number of data points (Formann, 1984). For the same reasons (also for considerations of missing data points and poor clinical feasibility/applicability), we did not include other biomarkers (i.e. CSF and imaging biomarkers) as clustering features in the main cluster solution; these were tested for *post hoc* comparisons.

Cluster analysis was performed in *R*. Agglomerative hierarchical clustering with Euclidean distance calculation was applied. Various measures were tested to estimate the optimal number of clusters. Of 24 solutions, 10 suggested two clusters, and seven suggested three clusters [see Supplementary Fig. 1 for the results from Hartigan's rule (value = 6.03 for three clusters)]. Ultimately, we prioritized the three-cluster solution because of a better balanced data distribution as well as clinical relevance (the two-cluster solution had 91% of patients in one cluster, limiting its utility). Furthermore, to check the stability of the clustering, we randomly half-split the study population and the same clustering method was repeated. This procedure was repeated 1000 times and the case memberships were cross-checked with that of the original clustering each time. Cohen's Kappa agreement rates were all in the moderate-substantial range (0.53 for mild motor-predominant, 0.54 for intermediate, and 0.70 for diffuse-malignant).

### Clinical definition of the subtypes

Cluster analysis is a statistical method that calculates and combines features at the group level. However, there will be individuals in each cluster who will not have all the key features that defined the cluster. For example, if a cluster solution states that RBD, orthostatic hypotension and cognition are the three most important variables for assigning to a cluster, there will still be patients in that cluster who do not have these variables. This means that results from clustering solutions cannot be applied to an individual, unless the solution is translated into rules to assign patients to clusters.

Therefore, based on analysis of characteristics of each cluster, we generated a categorical definition to assign individuals to the three subtypes. This was generated by identifying the critical features that discriminated clusters on principle component analysis (see 'Results' section, and Supplementary Fig. 2). On this analysis, motor markers (UPDRS II and III, PIGD score), autonomic dysfunction (SCOPA-AUT), RBD, and cognition were highly ranked significant features. This was converted into four composite domains: motor (UPDRS-II, UPDRS-III, and PIGD score), cognition (combining all available neuropsychological batteries in PPMI), RBD (RBDSQ score), and dysautonomia (SCOPA-AUT total score). Based on the distribution of scores in each cluster, the following definitions were created: Subtype I (mild motor-predominant): both composite motor score and all non-motor summary scores (NMS) are below the 75th percentile; Subtype III (diffuse malignant): EITHER (i) composite motor score $>$75th percentile and $\geqslant$1 of 3 non-motor scores $>$75th percentile; OR (ii) all three non-motor scores $>$75th percentile; and Subtype II (intermediate): those not meeting criteria for Subtype I or II

Further details are described in Supplementary Tables 1 and 2. Also to assist with clinical usefulness, we created an Excel 'Subtype Calculator' into which patient values can be entered, and the appropriate category automatically calculated (Supplementary material). Note that this can only be applied to patients who have similar characteristics to the PPMI population at baseline (e.g. *de novo* untreated Parkinson's disease).

### Principal component analysis

To assess the robustness of subtypes and the importance of each variable in separating them, we used principal component analysis (PCA). The loading value and *t*-score of each feature represent the importance of each variable in defining the subtypes. PCA model, loading values and $t_1$-$t_2$ scatter plot were created using *SIMCA* software (version 14.1) (MKS Data Analytics).

### Univariate comparisons

We evaluated differences in all baseline demographics, clinical characteristics, CSF biomarkers and imaging markers (including variables that were not used in cluster analysis) between subtypes/clusters. In addition, longitudinal changes in outcomes of interest and GCO were compared between subtypes. Univariate statistical tests were either one-way ANOVA (with Bonferroni *post hoc* test) or chi-square test where appropriate. Analysis of covariance (ANCOVA) for the between-clusters/subtypes comparisons was also applied, adjusting for age and disease duration.

### Structural MRI analysis

The Parkinson's disease-specific atrophy network (Zeighami *et al.*, 2015) was used as a region of interest to examine atrophy (compared to controls). An unpaired *t*-test was performed for each voxel within the predefined region of interest. The *P*-values were then corrected for multiple comparisons using the false discovery rate (FDR) technique. Furthermore, we also performed an exploratory voxel-wise uncorrected analysis to compare the whole brain atrophy pattern between clinical subtypes.

### General linear model

For analysis of progression, we applied general linear models (GLMs) for a more comprehensive longitudinal comparison between subtypes. In each separate GLM, the change in the outcome measure was defined as the dependent variable. In order to avoid 'regression towards mean' bias, multivariate statistical adjustment was performed using the baseline value of each outcome variable as a predictor (Vickers and Altman, 2001). Subtype membership was the main independent variable with follow-up duration as a potential covariate included in each model. All univariate and multivariate analyses were performed using IBM SPSS Statistics for Macintosh software (version 23.0). Two-tailed *P*-value $<0.05$ was considered the threshold for statistical significant differences in all analyses.

# Results

## Baseline description

A total of 421 *de novo* treatment-naive patients with Parkinson's disease were included in this study consisting of 276 (65.6%) males and 145 (34.4%) females with an average age of 61.1 $\pm$ 9.7 and Parkinson's disease duration

(defined in PPMI as time from motor symptom onset) of 6.5 ± 6.5 months at baseline. The mean MDS-UPDRS parts I, II and III were 5.6 ± 4.1, 6.0 ± 4.2 and 21.0 ± 9.0, respectively.

## Statistical clusters

Eighteen different variables were included in the final clustering solution: age, genetic risk score, orthostatic systolic blood pressure drop, MDS-UPDRS-Part II, MDS-UPDRS-Part III, tremor/PIGD scores, ESS, GDS, STAI, QUIP, RBDSQ, SCOPA-AUT, UPSIT, and average z-scores of visuospatial, speed/attention, memory and executive function (Supplementary Table 1).

As illustrated in Fig. 1, the cluster analysis revealed three distinct clusters of Parkinson's disease patients (Table 1 and Supplementary Fig. 1), with similar disease duration ($P = 0.23$). We observed substantial baseline differences in motor and non-motor manifestations between clusters, with clinically important effect sizes. Cluster I patients were younger and exhibited the mildest scores in motor UPDRS and sleep disorders, olfactory and autonomic dysfunction. They also had the least affected cognitive performance in almost all domains at baseline. We named this cluster 'mild motor-predominant'. By contrast, cluster III was the smallest group who presented with the worst scores in both motor and non-motor components (except for olfactory dysfunction and hallucinations). These patients had an average of MDS-UPDRS total score of 50.1 [standard deviation (SD) = 13.1] and the highest PIGD score at baseline. In addition, they also had the highest RBD score, Epworth score and worse autonomic function, the most severe cognitive impairment and most severe depressive, anxiety, apathy, pain and fatigue symptoms at baseline. We termed this cluster 'diffuse malignant'. The remaining cluster was categorized as class II and represented an intermediate status in most motor and non-motor manifestations at baseline. We named this cluster 'intermediate'. Except for UPSIT score (olfaction), all differences remained significant after adjustment for age and Parkinson's disease duration.

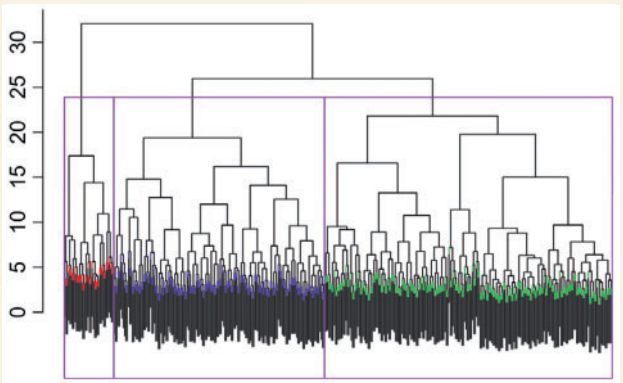


**Figure 1 Dendrogram of the final hierarchical cluster solution in the PPMI population.** Green = mild motor-predominant; blue = intermediate; red = diffuse malignant.

## Clinical subtypes

Following the exploration of clusters, clinical classification rules were developed to assign patients into discrete subtypes (see 'Materials and methods' section). Two hundred and twenty-three were categorized into the mild motor predominant subtype, 146 into intermediate, and 52 (12.4%) into diffuse malignant. The overall membership agreement rate between the statistical clusters and clinical subtypes was 76.3% for diffuse malignant, 74.2% for mild motor predominant, and 50.6% for intermediate. In general, differences in subtypes were even larger than for clusters (Table 2 and Supplementary Fig. 3). To summarize, the diffuse malignant subtype had the baseline highest MDS-UPDRS total score (51.7 ± 11.3), the highest PIGD score and the worst Schwab and England ADL, ESS, RBDSQ, GDS, STAI, UPSIT, SCOPA-AUT (including its subcomponents), apathy, fatigue and all cognitive scores (except HVLT-retention) (all $P < 0.05$). On the other side of the spectrum, the 'mild motor-predominant' subtype had the lowest severity of motor and non-motor manifestations with an average MDS-UPDRS total score half of subtype III (26.4 ± 9.4). These had the least impaired cognition, psychiatric features, sleep problems, olfactory and autonomic dysfunctions at baseline (all $P < 0.05$). For almost all manifestations, Parkinson's disease patients of the intermediate subtype ($n = 146$) had values intermediate between Subtypes I and III (Table 2). Most between-subtype differences remained statistically significant after additional adjustment for age and disease duration (Table 2).

## Importance of baseline clinical and biomarker features for subtype discrimination

The PCA model testing all clinical features, CSF, genetic and imaging biomarkers found that the strongest loading values (i.e. classification power) were for MDS-UPDRS-Part I (0.37), MDS-UPDRS-Part II (0.37), SCOPA-AUT (0.32), PIGD (0.28), RBDSQ (0.27), GDS (0.26), Schwab and England ADL (0.26), STAI (0.25), UPDRS-Part III (0.25), ESS (0.19) and composite cognitive status (0.18) (Fig. 2 and Supplementary Fig. 2).

## Imaging, blood and CSF biomarkers

We examined the imaging and CSF biomarkers in the clinical subtypes (Tables 1 and 2).

On MRI, the diffuse malignant subtype had a significantly lower Parkinson's disease-specific ICA network score (−0.33 ± 0.90, $P = 0.018$), indicating more severe atrophy in brain areas affected in early Parkinson's disease. Both the Parkinson's disease-specific ICA network and the substantia nigra demonstrated the least atrophy in the mild

**Table 1** ***Post hoc*** **comparison of the baseline demography and clinical features between the three statistical clusters of Parkinson's disease population in the PPMI based on hierarchical clustering solution (*n* = 421)**

| Characteristic | Cluster I Mild motor-predominant (*n* = 221) | Cluster II Intermediate (*n* = 162) | Cluster III Diffuse malignant (*n* = 38) | Unadjusted ANOVA/Chi$^2$ *P*-value | *Post hoc* Bonferroni *P*-value | Adjusted ANCOVA* *P*-value |
|---|---|---|---|---|---|---|
| **Baseline** | | | | | | |
| Age at onset (year) | 58.3 (9.9) | 65.4 (8.1) | 58.8 (8.9) | **<0.001** | **II versus I, II versus III** | - |
| Male sex (%) | 146 (66.1) | 107 (66.0) | 23 (60.5) | 0.791 | - | - |
| White race (%) | 214 (96.8) | 150 (92.6) | 35 (92.1) | 0.331 | - | - |
| Education history, years | 15.7 (2.8) | 15.6 (3.2) | 14.9 (3.1) | 0.363 | - | - |
| Symptoms duration, month | 6.1 (5.9) | 6.9 (7.0) | 7.7 (7.7) | 0.230 | - | - |
| Positive family history (%) | 35 (15.9) | 15 (9.4) | 5 (13.2) | 0.177 | - | - |
| Genetic Risk Score (GRS) | −0.015 (0.008) | −0.018 (0.009) | −0.016 (0.010) | **0.002** | **I versus II** | - |
| Pathologic variants (%) | | | | | | |
| *GBA* | 3 (1.5) | 4 (2.8) | 0 | 0.480 | - | - |
| *LRRK2* | 3 (1.5) | 3 (2.1) | 0 | 0.671 | - | - |
| UPDRS-total score | 26.1 (9.6) | 37.3 (12.1) | 50.1 (13.1) | **<0.001** | **All comparisons** | **<0.001** |
| **Motor symptoms and signs** | | | | | | |
| UPDRS-Part II | 4.1 (2.8) | 7.1 (4.0) | 12.1 (4.6) | **<0.001** | **All comparisons** | **<0.001** |
| UPDRS-Part III | 17.7 (7.5) | 24.1 (9.0) | 27.3 (9.4) | **<0.001** | **I versus II, I versus III** | **<0.001** |
| Schwab and England score | 94.5 (5.4) | 92.2 (5.9) | 89.1 (6.6) | **<0.001** | **All comparisons** | **<0.001** |
| Tremor score | 0.44 (0.28) | 0.55 (0.33) | 0.62 (0.41) | **<0.001** | **I versus II, I versus III** | **0.006** |
| PIGD score | 0.14 (0.14) | 0.30 (0.25) | 0.43 (0.28) | **<0.001** | **All comparisons** | **<0.001** |
| Tremor/PIGD phenotype (%) | | | | | | |
| Tremor-dominant | 180 (81.4) | 107 (66.0) | 21 (55.3) | | | |
| Intermediate | 15 (6.8) | 16 (9.9) | 6 (15.8) | **0.001** | - | - |
| PIGD-dominant | 26 (11.8) | 39 (24.1) | 11 (28.9) | | | |
| **Non-motor symptoms and signs** | | | | | | |
| UPDRS-Part I | 4.4 (3.1) | 6.0 (4.0) | 10.8 (4.9) | **<0.001** | **All comparisons** | **<0.001** |
| Hallucination | 0.03 (0.16) | 0.02 (0.16) | 0.08 (0.27) | 0.199 | - | 0.223 |
| Apathy | 0.14 (0.36) | 0.19 (0.50) | 0.61 (0.79) | **<0.001** | **III versus I, II** | **<0.001** |
| Pain | 0.60 (0.77) | 0.73 (0.81) | 1.29 (1.14) | **<0.001** | **III versus I, II** | **<0.001** |
| Fatigue | 0.47 (0.61) | 0.75 (0.81) | 1.34 (1.12) | **<0.001** | **All comparisons** | **<0.001** |
| Epworth sleepiness score | 5.1 (2.8) | 6.2 (3.8) | 7.8 (4.3) | **<0.001** | **All comparisons** | **<0.001** |
| GDS | 2.1 (2.4) | 2.0 (1.7) | 5.3 (3.3) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| STAI score | 64.4 (17.7) | 62.1 (16.0) | 86.7 (18.7) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| Impulse control disorders (QUIP sum score) | 0.28 (0.52) | 0.08 (2.73) | 0.97 (1.13) | **<0.001** | **All comparisons** | **<0.001** |
| RBD score | 2.7 (2.2) | 4.0 (2.9) | 7.1 (2.5) | **<0.001** | **All comparisons** | **<0.001** |
| Olfaction UPSIT percentile | 23.6 (8.0) | 20.4 (8.3) | 23.0 (7.9) | **0.001** | **I versus II** | 0.221 |
| Drop in systolic blood pressure (mmHg) | 2.5 (11.1) | 7.0 (13.7) | 6.0 (14.2) | **0.002** | **I versus II** | **0.013** |
| SCOPA autonomic questionnaire | | | | | | |
| Orofacial symptoms | 0.8 (1.1) | 1.1 (1.2) | 2.7 (1.7) | **<0.001** | **All comparisons** | **<0.001** |
| Constipation | 0.8 (1.1) | 1.3 (1.5) | 1.8 (1.4) | **<0.001** | **I versus II, I versus III** | **<0.001** |
| Urinary | 3.5 (2.4) | 4.9 (4.9) | 7.5 (8.4) | **<0.001** | **All comparisons** | **<0.001** |

(continued)

**Table 1 Continued**

| Characteristic | Cluster I Mild motor-predominant (*n* = 221) | Cluster II Intermediate (*n* = 162) | Cluster III Diffuse malignant (*n* = 38) | Unadjusted ANOVA/Chi$^2$ *P*-value | *Post hoc* Bonferroni *P*-value | Adjusted ANCOVA* *P*-value |
|---|---|---|---|---|---|---|
| Cardiovascular | 0.4 (0.6) | 0.4 (0.7) | 1.2 (1.4) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| Thermoregulatory | 1.0 (1.2) | 1.1 (1.2) | 2.6 (2.2) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| Pupillomotor | 0.3 (0.6) | 0.4 (0.6) | 0.8 (0.9) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| Sexual | 0.9 (1.3) | 1.3 (1.6) | 1.9 (2.1) | **<0.001** | **I versus II, I versus III** | **0.001** |
| Total score | 7.6 (4.6) | 10.5 (6.0) | 17.6 (7.6) | **<0.001** | **All comparisons** | **<0.001** |
| Cognitive function | | | | | | |
| MOCA (adjusted score) | 27.4 (2.3) | 27.0 (2.2) | 26.8 (2.6) | 0.170 | - | 0.570 |
| Benton judgment of line orientation (adjusted score) | 13.0 (2.4) | 11.0 (3.1) | 10.7 (2.8) | **<0.001** | **I versus II, I versus III** | **<0.001** |
| HVLT-total recall (T-score) | 47.2 (11.5) | 46.3 (10.6) | 42.6 (9.7) | 0.058 | - | 0.056 |
| HVLT-delayed recall (T-score) | 48.0 (11.4) | 46.4 (11.7) | 42.7 (13.4) | **0.029** | **I versus III** | **0.032** |
| HVLT-retention (T-score) | 50.2 (11.0) | 49.8 (11.9) | 47.4 (14.8) | 0.407 | - | 0.451 |
| HVLT-recognition discrimination index (T-score) | 49.2 (10.8) | 49.5 (13.2) | 48.6 (15.9) | 0.921 | - | 0.945 |
| LNS scaled score | 11.8 (2.7) | 11.3 (2.7) | 10.6 (2.6) | **0.015** | **I versus III** | **0.016** |
| Semantic fluency (T-score) | 51.6 (10.1) | 50.4 (9.3) | 48.0 (10.4) | 0.095 | - | 0.058 |
| Symbol digit test (T-score) | 45.6 (8.0) | 45.7 (8.8) | 39.2 (9.4) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| **Brain imaging** | | | | | | |
| SPECT SBR | | | | | | |
| Caudate | 2.07 (0.54) | 1.85 (0.50) | 2.09 (0.83) | **0.012** | **I versus II** | 0.085 |
| Putamen | 0.18 (0.85) | −0.19 (0.76) | 0.25 (1.79) | **0.012** | **I versus II** | **0.033** |
| Delay between first clinical evaluation and SPECT acquisition (days) | 13 (45) | 18 (42) | 12 (32) | 0.478 | - | - |
| MRI (DBM) | | | | | | |
| Parkinson's disease-related ICA network | 0.23 (0.97) | −0.26 (0.97) | −0.08 (1.1) | **0.002** | **I versus II** | 0.091 |
| Substantia nigra score | 0.11 (1.06) | −0.13 (0.96) | −0.02 (0.81) | 0.232 | - | 0.636 |
| Delay between first clinical evaluation and MRI acquisition (days) | 33 (30) | 30 (27) | 25 (20) | 0.338 | - | - |
| **CSF biomarkers** | | | | | | |
| α-Synuclein | 1800.0 (746.2) | 1936.5 (861.1) | 1741.0 (727.9) | 0.180 | - | 0.625 |
| Aβ42 | 377.1 (100.1) | 368.0 (95.6) | 342.2 (114.7) | 0.137 | - | 0.143 |
| P-tau | 16.3 (10.2) | 15.5 (10.6) | 13.9 (7.7) | 0.404 | - | 0.321 |
| T-tau | 42.5 (16.0) | 48.9 (21.0) | 42.1 (18.1) | **0.003** | **I versus II** | 0.351 |
| Aβ42/T-tau ratio | 0.24 (0.11) | 0.21 (0.08) | 0.23 (0.12) | 0.070 | - | 0.285 |
| Aβ42/α-synuclein ratio | 0.010 (0.007) | 0.009 (0.009) | 0.010 (0.010) | 0.534 | - | 0.532 |
| P-tau/α-synuclein ratio | 1800.0 (746.2) | 1936.5 (861.1) | 1741.0 (727.9) | 0.180 | - | 0.575 |
| Delay between first clinical evaluation and LP acquisition (days) | 34 (29) | 31 (25) | 27 (21) | 0.218 | - | - |

*ANCOVA was performed to adjust between-clusters comparisons for baseline age and duration of disease as potential covariates.

Aβ = amyloid-β; DBM = deformation-based morphometry; GDS = Geriatric depression scale; LNS = Letter number sequencing; LP = lumbar puncture; SBR = striatal binding ratios; STAI = State-trait anxiety inventory; UPSIT = University of Pennsylvania Smell Identification Test.

**Table 2** ***Post hoc*** **comparison of the baseline demography and clinical features, brain imaging and CSF biomarkers between the three subtypes of Parkinson's disease population in the PPMI based on the recommended clinical definition (*n* = 421)**

| Characteristic | Phenotype I Mild motor-predominant (*n* = 223) | Phenotype II Intermediate (*n* = 146) | Phenotype III Diffuse malignant (*n* = 52) | Unadjusted ANOVA/Chi$^2$ *P*-value | *Post hoc* Bonferroni *P*-value | Adjusted ANCOVA* *P*-value |
|---|---|---|---|---|---|---|
| **Baseline** | | | | | | |
| Age at onset, years | 60.0 (9.7) | 62.2 (9.6) | 62.8 (9.6) | **0.038** | **None** | - |
| Male sex (%) | 140 (62.8) | 98 (67.1) | 38 (73.1) | 0.329 | - | - |
| White race (%) | 215 (96.4) | 135 (92.5) | 49 (94.2) | 0.331 | - | - |
| Education history, year | 15.7 (3.0) | 15.4 (2.9) | 15.4 (3.2) | 0.514 | - | - |
| Symptoms duration, month | 6.6 (6.7) | 6.2 (6.2) | 7.4 (7.0) | 0.531 | - | - |
| Positive family history (%) | 29 (13.1) | 18 (12.5) | 8 (15.4) | 0.869 | - | - |
| Genetic risk score | −0.015 (0.008) | −0.17 (0.010) | −0.018 (0.009) | 0.170 | - | - |
| Pathologic variants (%) | | | | | | |
| *GBA* | 3 (1.5) | 2 (1.5) | 2 (4.3) | 0.401 | - | - |
| *LRRK2* | 3 (1.5) | 3 (2.2) | 0 | 0.570 | - | - |
| UPDRS total score | 26.4 (9.4) | 35.2 (11.8) | 51.7 (11.3) | **<0.001** | **All comparisons** | **<0.001** |
| **Motor symptoms and signs** | | | | | | |
| UPDRS-Part II | 4.0 (2.7) | 6.9 (3.9) | 11.8 (4.3) | **<0.001** | **All comparisons** | **<0.001** |
| UPDRS-Part III | 18.3 (7.5) | 21.9 (8.9) | 30.0 (9.2) | **<0.001** | **All comparisons** | **<0.001** |
| Schwab and England score | 94.7 (5.0) | 92.2 (6.2) | 88.9 (6.5) | **<0.001** | **All comparisons** | **<0.001** |
| Tremor score | 0.47 (0.32) | 0.50 (0.30) | 0.58 (0.38) | 0.085 | - | 0.334 |
| PIGD score | 0.15 (0.16) | 0.26 (0.23) | 0.46 (0.28) | **<0.001** | **All comparisons** | **<0.001** |
| Tremor/PIGD phenotype (%) | | | | | | |
| Tremor-dominant | 177 (79.4) | 104 (71.2) | 27 (51.9) | | | |
| Intermediate | 10 (4.5) | 20 (13.7) | 7 (13.5) | **<0.001** | **-** | **-** |
| PIGD-dominant | 36 (16.1) | 22 (15.1) | 18 (34.6) | | | |
| **Non-motor symptoms and signs** | | | | | | |
| UPDRS-Part I | 4.0 (3.0) | 6.5 (3.6) | 9.9 (5.3) | **<0.001** | **All comparisons** | **<0.001** |
| Hallucination | 0.03 (0.16) | 0.03 (0.16) | 0.06 (0.24) | 0.492 | - | 0.452 |
| Apathy | 0.14 (0.38) | 0.16 (0.39) | 0.54 (0.87) | **<0.001** | **III versus I, II** | **<0.001** |
| Pain | 0.57 (0.74) | 0.85 (0.91) | 0.98 (0.96) | **<0.001** | **I versus II, III** | **<0.001** |
| Fatigue | 0.44 (0.61) | 0.74 (0.74) | 1.33 (1.12) | **<0.001** | **All comparisons** | **<0.001** |
| Epworth sleepiness score | 5.1 (3.0) | 6.1 (3.5) | 7.6 (4.1) | **<0.001** | **All comparisons** | **<0.001** |
| GDS | 1.9 (2.2) | 2.4 (2.3) | 4.1 (3.0) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| STAI score | 62.2 (16.0) | 65.9 (18.1) | 78.7 (22.7) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| Impulse control disorders (QUIP sum score) | 0.22 (0.55) | 0.28 (0.56) | 0.40 (0.77) | 0.111 | - | **0.049** |
| RBD score | 2.1 (1.5) | 4.6 (2.9) | 6.6 (3.0) | **<0.001** | **All comparisons** | **<0.001** |
| Olfaction UPSIT percentile | 23.0 (8.0) | 22.1 (8.1) | 19.8 (9.3) | **0.040** | **I versus III** | 0.145 |
| Drop in systolic blood pressure (mmHg) | 3.3 (12.0) | 5.1 (11.3) | 8.3 (17.4) | **0.030** | **I versus III** | **0.044** |
| SCOPA autonomic questionnaire | | | | | | |
| Orofacial symptoms | 0.6 (0.9) | 1.3 (1.3) | 2.5 (1.6) | **<0.001** | **All comparisons** | **<0.001** |
| Constipation | 0.6 (0.9) | 1.5 (1.5) | 2.0 (1.5) | **<0.001** | **All comparisons** | **<0.001** |
| Urinary | 3.2 (2.0) | 5.0 (3.2) | 7.9 (9.9) | **<0.001** | **All comparisons** | **<0.001** |

(continued)

**Table 2** Continued

| Characteristic | Phenotype I Mild motor-predominant (*n* = 223) | Phenotype II Intermediate (*n* = 146) | Phenotype III Diffuse malignant (*n* = 52) | Unadjusted ANOVA/Chi$^2$ *P*-value | *Post hoc* Bonferroni *P*-value | Adjusted ANCOVA* *P*-value |
|---|---|---|---|---|---|---|
| Cardiovascular | 0.3 (0.5) | 0.6 (0.9) | 0.9 (1.1) | **<0.001** | **All comparisons** | **<0.001** |
| Thermoregulatory | 0.8 (1.1) | 1.4 (1.4) | 2.0 (1.8) | **<0.001** | **All comparisons** | **<0.001** |
| Pupillomotor | 0.3 (0.6) | 0.5 (0.6) | 0.8 (0.8) | **<0.001** | **III versus I, III versus II** | **<0.001** |
| Sexual | 0.8 (1.2) | 1.3 (1.7) | 1.7 (2.0) | **<0.001** | **I versus II, I versus III** | **<0.001** |
| Total score | 6.6 (3.4) | 11.7 (6.5) | 16.5 (6.6) | **<0.001** | **All comparisons** | **<0.001** |
| Cognitive function | | | | | | |
| MoCA (adjusted score) | 27.5 (2.1) | 26.8 (2.5) | 26.7 (2.5) | **0.005** | **I versus II, I versus III** | **0.020** |
| Benton judgment of line orientation (adjusted score) | 12.6 (2.6) | 11.7 (3.2) | 10.5 (2.9) | **<0.001** | **All comparisons** | **<0.001** |
| HVLT-total recall (T-score) | 48.0 (11.0) | 45.4 (11.1) | 42.9 (10.2) | **0.004** | **I versus III** | **0.004** |
| HVLT-delayed recall (T-score) | 48.3 (11.4) | 46.4 (12.1) | 42.3 (11.1) | **0.003** | **I versus III** | **0.004** |
| HVLT-retention (T score) | 50.9 (11.6) | 49.0 (11.2) | 48.1 (13.3) | 0.093 | - | 0.131 |
| HVLT-recognition discrimination index (T-score) | 50.4 (11.3) | 47.1 (11.8) | 50.2 (16.3) | **0.039** | **I versus II** | **0.040** |
| LNS scaled score | 12.0 (2.7) | 11.0 (2.6) | 10.8 (2.6) | **0.001** | **I versus II, I versus III** | **0.002** |
| Semantic fluency (T-score) | 52.1 (9.6) | 49.9 (9.6) | 47.6 (10.8) | **0.004** | **I versus III** | **0.002** |
| Symbol digit test (T-score) | 46.9 (7.6) | 43.8 (9.0) | 41.2 (9.7) | **<0.001** | **I versus II, I versus III** | **<0.001** |
| **Brain imaging** | | | | | | |
| SPECT (SBR) | | | | | | |
| Caudate | 2.10 (0.53) | 1.90 (0.52) | 1.73 (0.71) | **0.001** | **I versus II, I versus III** | **0.002** |
| Putamen | 0.23 (0.81) | −0.18 (0.91) | −0.14 (1.40) | **0.006** | **I versus II** | **0.006** |
| Delay between first clinical evaluation and SPECT acquisition (days) | 19 (42) | 11 (48) | 10 (26) | 0.178 | - | - |
| MRI (DBM) | | | | | | |
| PD-related ICA network | 0.17 (1.04) | −0.13 (0.94) | −0.33 (0.90) | **0.018** | **I versus III** | **0.022** |
| Substantia nigra score | 0.10 (1.09) | −0.13 (0.91) | −0.03 (0.86) | 0.269 | - | 0.235 |
| Delay between first clinical evaluation and MRI acquisition (days) | 33 (32) | 30 (25) | 26 (19) | 0.267 | - | - |
| **CSF biomarkers** | | | | | | |
| α-synuclein | 1823.3 (756.7) | 1890.0 (863.4) | 1829.7 (741.3) | 0.727 | - | 0.842 |
| Aβ42 | 378.3 (97.7) | 373.3 (101.8) | 329.0 (96.7) | **0.006** | **III versus I, III versus II** | **0.009** |
| P-tau | 16.2 (11.3) | 15.5 (9.3) | 14.3 (6.6) | 0.458 | - | 0.392 |
| T-tau | 43.7 (16.7) | 46.2 (20.0) | 46.6 (21.6) | 0.385 | - | 0.830 |
| Aβ42/t-tau ratio | 9.5 (3.1) | 9.1 (3.2) | 8.2 (3.0) | **0.032** | **I versus III** | 0.136 |
| Aβ42/α-synuclein ratio | 0.24 (0.10) | 0.23 (0.09) | 0.21 (0.10) | 0.223 | - | 0.424 |
| P-tau/α-synuclein ratio | 0.010 (0.009) | 0.009 (0.006) | 0.010 (0.008) | 0.583 | - | 0.597 |
| Delay between first clinical evaluation and LP acquisition, days | 34 (30) | 32 (24) | 28 (18) | 0.321 | - | - |

*ANCOVA was performed to adjust between-subtypes comparisons for baseline age and duration of disease as potential covariates. All values in brackets are standard deviation (SD) unless otherwise specified.
Aβ = amyloid-β; DBM = deformation-based morphometry; GDS = Geriatric Depression Scale; LNS = letter number sequencing; LP = lumbar puncture; SBR = striatal binding ratios; STAI = state-trait anxiety inventory; UPSIT = University of Pennsylvania Smell Identification Test.

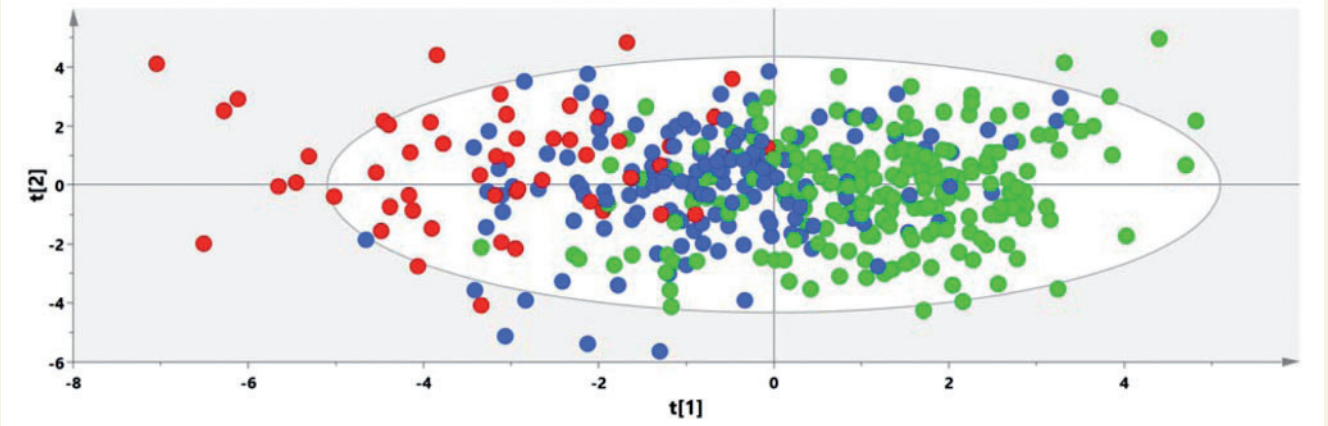


**Figure 2 Score scatter plot of the PCA to discriminate different clinical phenotype in the PPMI population.** Green = mild motor-predominant; blue = intermediate; red = diffuse malignant.

motor-predominant subtype. To illustrate the global atrophy pattern in different Parkinson's disease clinical subtypes, the DBM brain atrophy measure was compared with controls and between subtypes. For all three subtypes, midbrain regions midbrain (more specifically, substantia nigra) were significantly atrophied. The total volume of regions significantly different in atrophy in patients than controls increased in a stepwise manner from subtype I (0.33 $cm^3$), to II (1.45 $cm^3$), to III (1.74 $cm^3$). Figure 3A shows the corresponding FDR-corrected maps for each subtype. The effect size *t*-maps (Fig. 3B) demonstrated a trend for progressively greater whole brain atrophy in each subtype compared to controls. Figure 3D and E illustrate the exploratory voxel-wise comparisons between subtypes. Greater atrophy was seen in the diffuse malignant subtype compared to both mild-motor predominant and intermediate subtypes, in a stepwise manner, although no voxel differences survived statistical correction for multiple comparisons.

On dopaminergic SPECT scanning, the mild-motor predominant subtype had the least denervation of both caudate ($2.10 \pm 0.53$, $P = 0.001$) and putamen ($0.23 \pm 0.81$, $P = 0.006$). By contrast, the 'diffuse malignant' subtype had the highest level of caudate denervation ($1.73 \pm 0.71$, $P = 0.001$).

On CSF analysis, there was lower CSF amyloid-β ($329.0 \pm 96.7$ pg/ml, $P = 0.006$) and amyloid-β/t-tau ratio ($8.2 \pm 3.0$, $P = 0.032$) in the diffuse malignant subtype. The mild motor-predominant subtype had the highest CSF amyloid-β ($378.3 \pm 97.7$ pg/ml) levels and amyloid-β/t-tau ratio ($9.5 \pm 3.1$) (all $P < 0.05$). There were no differences in CSF α-synuclein between subtypes ($P = 0.73$), and no difference in any blood-based biomarker (data not shown).

In contrast to the clinical subtypes, differences between the statistical clusters failed to reach statistical significance for biomarkers, suggesting that the clinical subtypes were able to identify more robust differences than the statistical clusters.

## Disease progression

The PPMI population was followed for $32.8 \pm 9.3$ months, with no difference in follow-up duration between subtypes (Table 3 and Fig. 4). Results from the GLM adjusted for baseline values and follow-up duration time showed that the intermediate and diffuse malignant subtypes had significantly greater progression in UPDRS-Part II (2.1 and 3.4 units more increase in compared to the mild motor-predominant subtype) and Schwab and England ADL score (2.4 and 6.9% more decline, respectively). Similar hierarchical progression could be also seen in several non-motor features namely UPDRS-Part I (1.7 points faster in intermediate, 2.7 points faster in diffuse malignant), ESS (0.9 and 1.7 points faster), STAI (3.2 and 8.7 units faster) and MoCA (0.7 and 1.5 points faster). Moreover, the speed of progression in the global composite outcome was significantly higher in the diffuse malignant (0.68 increase in z-score, $P < 0.001$) and intermediate (0.38 increase in z-score, $P < 0.001$) subtypes. As illustrated in Fig. 4, the faster slope of progression in the diffuse malignant subtype was most noticeable for Schwab and England ADL, MoCA, and GCO. When the GCO was compared between classically-defined tremor/PIGD phenotypes, those classified as PIGD did not have significantly worse prognosis (0.19 increase in z-score, $P = 0.087$). When looking at statistical clusters, results were similar in direction to the clinical subtypes, but effect sizes were larger for the clinical subtypes than the statistical clusters (Supplementary Fig. 4 and Supplementary Table 3).

On biomarker analysis patients with Parkinson's disease in the diffuse malignant subtype had greater decline in dopaminergic innervation of the caudate (0.15 units more decline, $P = 0.007$) and putamen (0.08 units more decline, $P = 0.006$) after an average of 2.7 years of follow-up.

As per definition, there were two ways to make a classificaiton of the diffuse malignant subtype; a minority ($n = 5$) fulfilled only the second criterion (high severity in all non-

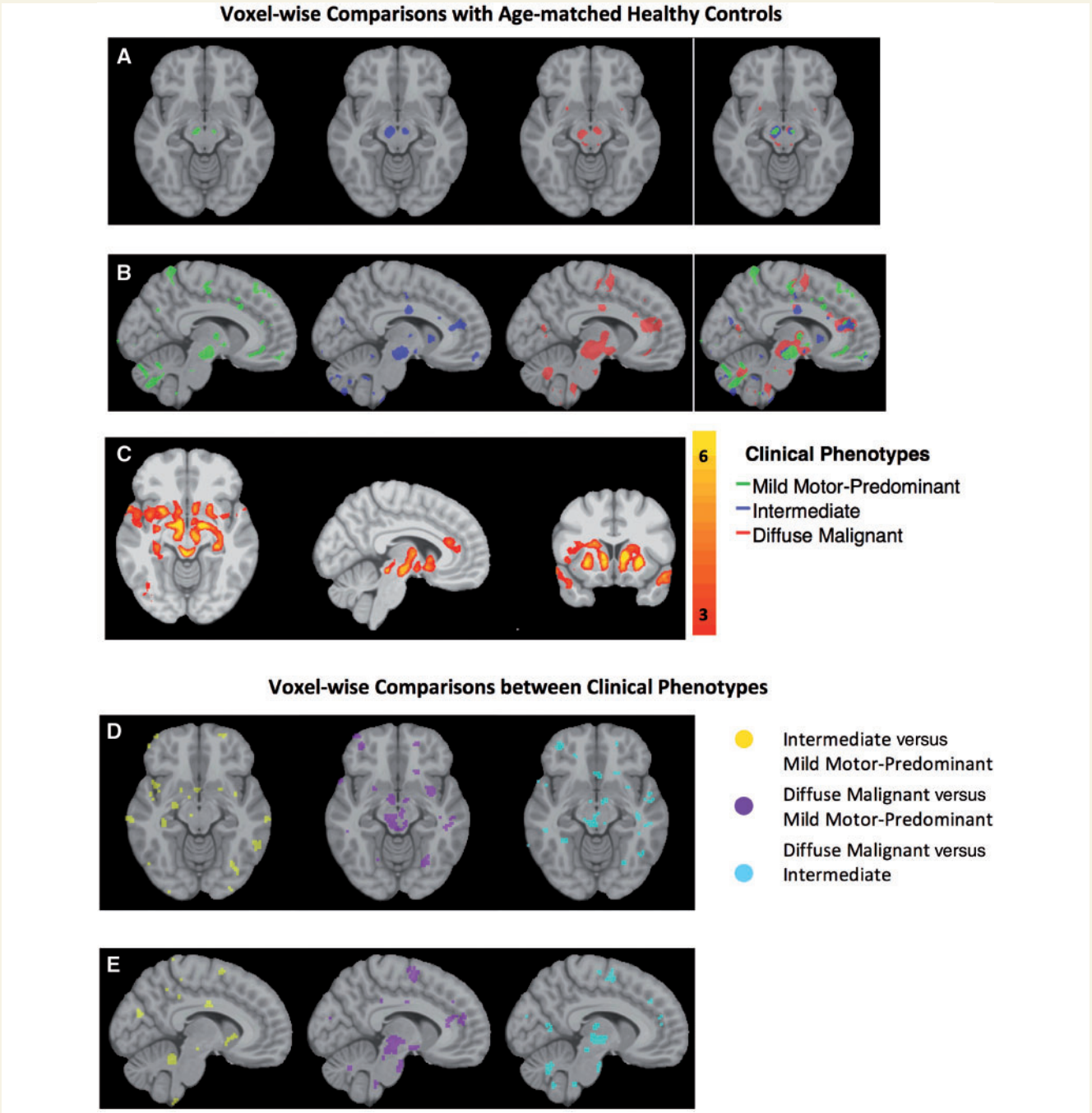


**Figure 3 Structural MRI analysis in different clinical phenotypes of Parkinson's disease in the PPMI population.** (**A**) DBM maps within the Parkinson's disease-specific network have been compared between each subtype and healthy controls, corrected for multiple comparison using FDR. There is higher atrophy in the diffuse malignant subtype compared to the intermediate subtype, and in the intermediate subtype compared to the mild motor-predominant subtype. (**B**) Whole brain atrophy pattern within each group compared to controls (uncorrected) shows consistent increase in the amount of atrophy from the mild motor-predominant to diffuse malignant subtype. The last column on the *right* shows the overlapped pattern of atrophy in all subtypes merged together. (**C**) DBM maps of the Parkinson's disease-specific network showing significant differences in atrophy between patients with Parkinson's disease and healthy control subjects in PPMI ($P = 0.003$ after Bonferroni correction for multiple comparisons). The independent component analysis spatial map was converted to z-statistic image via a normalized mixture-model fit and then thresholded at z = 3. Selected sections in MNI space at coordinates: $z = -10$, $x = -6$, $y = +14$. (**D** and **E**) Exploratory comparisons of the whole brain atrophy pattern between clinical subtypes (uncorrected) shows brain areas with higher atrophy in the intermediate subtype compared with the mild motor-predominant subtype (*left*), brain areas with higher atrophy in the diffuse malignant subtype compared with the mild motor-predominant (*middle*) and intermediate (*right*) subtypes.

**Table 3 Longitudinal changes in clinical motor and non-motor outcomes and SPECT striatal binding ratios in three different clinical phenotypes of Parkinson's disease population in the PPMI population with at least 1 year of follow-up ($n$ = 401)**

| Outcome | Phenotype I Mild motor-predominant ($n$ = 216) | Phenotype II Intermediate ($n$ = 136) | Phenotype III Diffuse malignant ($n$ = 49) | ANOVA $P$-value | Multivariate linear regression $P$-value |
|---|---|---|---|---|---|
| Follow-up time, months | 32.6 (9.0) | 32.8 (9.7) | 33.5 (9.7) | 0.825 | - |
| UPDRS-Part II | | | | | |
| Δ ($t_2$-$t_1$) | 2.6 (4.6) | 4.0 (5.5) | 3.6 (6.7) | **0.049** | - |
| β adjusted coefficient (95% CI) | 0[a] | 2.1 (1.0–3.2) | 3.4 (1.6–5.1) | - | **$P_{phenotype\ II} < 0.001$** **$P_{phenotype\ III} < 0.001$** |
| UPDRS-Part III | | | | | |
| Δ ($t_2$-$t_1$) | 4.8 (10.0) | 6.4 (12.0) | 3.1 (12.2) | 0.141 | - |
| β adjusted coefficient (95% CI) | 0[a] | 2.7 (0.5–4.9) | 2.1 (−1.3–5.4) | - | **$P_{phenotype\ II} = 0.015$** $P_{phenotype\ III} = 0.231$ |
| Schwab and England ADL score | | | | | |
| Δ ($t_2$-$t_1$) | −5.1 (8.9) | −6.7 (9.0) | −9.0 (11.7) | **0.023** | - |
| β adjusted coefficient (95% CI) | 0[a] | −2.4 (−4.2–−0.6) | −6.9 (−9.5–−4.2) | - | **$P_{phenotype\ II} = 0.008$** **$P_{phenotype\ III} < 0.001$** |
| Follow-up time, months | 28.4 (9.7) | 28.7 (10.4) | 29.5 (11.0) | 0.780 | - |
| UPDRS-Part I | | | | | |
| Δ ($t_2$-$t_1$) | 2.2 (4.0) | 3.0 (5.3) | 2.5 (6.1) | 0.283 | - |
| β adjusted coefficient (95% CI) | 0[a] | 1.7 (0.7–2.7) | 2.7 (1.2–4.3) | - | **$P_{phenotype\ II} = 0.001$** **$P_{phenotype\ III} = 0.001$** |
| Epworth sleepiness score | | | | | |
| Δ ($t_2$-$t_1$) | 0.9 (3.5) | 1.4 (4.2) | 1.6 (4.5) | 0.337 | - |
| β adjusted coefficient (95% CI) | 0[a] | 0.9 (0.1–1.7) | 1.7 (0.5–2.9) | - | **$P_{phenotype\ II} = 0.023$** **$P_{phenotype\ III} = 0.005$** |
| RBD score | | | | | |
| Δ ($t_2$-$t_1$) | 0.7 (2.1) | 0.4 (2.6) | −0.7 (2.9) | **0.001** | - |
| β adjusted coefficient (95% CI) | 0[a] | 0.5 (0.0–1.1) | 0.1 (−0.7–0.9) | - | **$P_{phenotype\ II} = 0.049$** $P_{phenotype\ III} = 0.761$ |
| SCOPA-AUT score | | | | | |
| Δ ($t_2$-$t_1$) | 2.7 (4.1) | 2.3 (6.1) | 1.0 (6.6) | 0.102 | - |
| β adjusted coefficient (95% CI) | 0[a] | 1.3 (0.1–2.4) | 1.4 (−0.4–3.1) | - | **$P_{phenotype\ II} = 0.030$** $P_{phenotype\ III} = 0.127$ |
| MoCA score | | | | | |
| Δ ($t_2$-$t_1$) | −0.6 (2.5) | −1.1 (3.1) | −1.8 (2.9) | **0.019** | - |
| β adjusted coefficient (95% CI) | 0[a] | −0.7 (−1.3–−0.2) | −1.5 (−2.3–−0.6) | - | **$P_{phenotype\ II} = 0.011$** **$P_{phenotype\ III} = 0.001$** |
| Geriatric depression scale | | | | | |
| Δ ($t_2$-$t_1$) | 0.3 (2.5) | 0.5 (2.8) | 0.7 (2.4) | 0.436 | - |
| β adjusted coefficient (95% CI) | 0[a] | 0.4 (−0.1–0.9) | 1.3 (0.5–2.1) | - | $P_{phenotype\ II} = 0.128$ **$P_{phenotype\ III} = 0.001$** |
| State-trait anxiety inventory | | | | | |
| Δ ($t_2$-$t_1$) | −0.9 (14.8) | 1.0 (14.7) | 1.4 (19.8) | 0.429 | - |
| β adjusted coefficient (95% CI) | 0[a] | 3.2 (0.3–6.2) | 8.7 (4.2–13.1) | - | **$P_{phenotype\ II} = 0.031$** **$P_{phenotype\ III} < 0.001$** |
| QUIP score | | | | | |
| Δ ($t_2$-$t_1$) | 0.0 (1.1) | 0.0 (0.9) | −0.1 (1.6) | 0.724 | - |
| β adjusted coefficient (95% CI) | 0[a] | 0.0 (−0.2–0.2) | 0.1 (−0.2–0.3) | - | $P_{phenotype\ II} = 0.836$ $P_{phenotype\ III} = 0.663$ |
| z-score | | | | | |
| Δ ($t_2$-$t_1$) | 0.56 (0.77) | 0.80 (0.89) | 0.82 (0.98) | **0.015** | - |
| β adjusted coefficient (95% CI) | 0[a] | 0.38 (0.20–0.56) | 0.68 (0.38–0.97) | - | **$P_{phenotype\ II} < 0.001$** **$P_{phenotype\ III} < 0.001$** |
| Right caudate | | | | | |
| Δ ($t_2$-$t_1$) | −0.28 (0.33) | −0.30 (0.38) | −0.34 (0.32) | 0.527 | - |
| β adjusted coefficient (95% CI) | 0[a] | −0.06 (−0.13–0.01) | −0.15 (−0.26–−0.04) | - | $P_{phenotype\ II} = 0.105$ **$P_{phenotype\ III} = 0.007$** |

(continued)

**Table 3 Continued**

| Outcome | Phenotype I Mild motor-predominant ($n$ = 216) | Phenotype II Intermediate ($n$ = 136) | Phenotype III Diffuse malignant ($n$ = 49) | ANOVA $P$-value | Multivariate linear regression $P$-value |
|---|---|---|---|---|---|
| Left caudate | | | | | |
| Δ ($t_2$-$t_1$) | −0.27 (0.31) | −0.27 (0.34) | −0.28 (0.33) | 0.985 | - |
| β adjusted coefficient (95% CI) | 0[a] | −0.04 (−0.11–0.02) | −0.06 (−0.16–0.04) | - | $P_{phenotype\ II}$ = 0.202<br>$P_{phenotype\ III}$ = 0.275 |
| Right putamen | | | | | |
| Δ ($t_2$-$t_1$) | −0.17 (0.21) | −0.12 (0.23) | −0.15 (0.21) | 0.125 | - |
| β adjusted coefficient (95% CI) | 0[a] | 0.02 (−0.03–0.06) | −0.03 (−0.09–0.04) | - | $P_{phenotype\ II}$ = 0.456<br>$P_{phenotype\ III}$ = 0.402 |
| Left putamen | | | | | |
| Δ ($t_2$-$t_1$) | −0.15 (0.20) | −0.11 (0.21) | −0.20 (0.27) | 0.065 | - |
| β adjusted coefficient (95% CI) | 0[a] | −0.01 (−0.05–0.02) | −0.08 (−0.13–−0.02) | - | $P_{phenotype\ II}$ = 0.461<br>**$P_{phenotype\ III}$ = 0.006** |

All presented values are mean (standard deviation), unless otherwise specified.
Statistical adjustment was performed using follow-up duration and baseline value of the outcome variable as potential covariates in the multivariate linear regression model to compare the amount of change in each outcome of interest between three phenotypes.
[a]Reference group.

motor domains with relatively preserved motor scores). This 'all-non-motor' subgroup demonstrated the worst prognosis of all groups. Compared to the rest of the diffuse malignant subtype, they had faster progression of GCO ($1.9 \pm 1.3$ versus $0.7 \pm 0.9$, $P = 0.006$) motor disability (MDS-UPDRS-Part III) ($11.4 \pm 9.4$ versus $2.1 \pm 12.2$, $P = 0.106$), Schwab and England ($20.0 \pm 12.2$ versus $7.7 \pm 11.1$, $P = 0.025$) and MoCA ($5.5 \pm 5.2$ versus $1.4 \pm 2.5$, $P = 0.076$).

# Discussion

Capitalizing on the large sample size and extensive assessment of clinical features and biomarkers in the PPMI, we were able to systematically identify subtypes of Parkinson's disease and their prognosis. Based predominantly on clinical features, we defined three Parkinson's disease subtypes. Our results showed that, despite similar disease duration, subtypes differed substantially in terms of neuroimaging, CSF biomarkers, clinical characteristics and disease progression over time.

## Clinical features

The variables we tested to create Parkinson's disease clusters and used for *post hoc* comparisons are the most complete set used so far, including not only a comprehensive clinical assessment but also neuroimaging, genetic, blood, and CSF biomarkers. The three clusters differed in most motor and non-motor features at the baseline evaluation. Moreover, to overcome practical problems with the application of clustering to the individual level, we created a clinical definition for subtyping Parkinson's disease at baseline, which can be applied in real-life practice. Note that the choice of clinical criteria was primarily based on clustering and PCA analyses, but also influenced by knowledge of the literature and other considerations. For instance, we elected not to use depression/anxiety (one of the significant features for cluster discrimination in PCA analysis) in the clinical subtyping definition since many psychiatric scales have limited content validity, are affected by large ceiling effects, and can be confounded by non-mood symptoms such as sleep disturbances and motor impairment (already included in the clinical subtyping definition) (Schrag *et al.*, 2007). In line with previous studies (Fereshtehnejad *et al.*, 2015), four main variables—MDS-UPDRS, RBD, autonomic dysfunction, and cognitive impairment—were the most critical clinical determinants of prognosis.

*Post hoc* comparisons of other clinical variables and especially analysis of neuroimaging and CSF biomarker profiles demonstrated that the clinical subtypes represent distinct subtypes of Parkinson's disease, which have different rates of progression. The diffuse malignant subtype consisted of 12% of the Parkinson's disease sample and had the most severe motor and non-motor features, suggesting simultaneous involvement of dopaminergic and non-dopaminergic pathways at baseline. The diffuse malignant cluster was the most robustly identified; for example, in two cluster solutions, a highly-similar cluster was clearly identified, with considerable overlap in membership to the three-cluster solution diffuse malignant cluster (data not shown). Even after only 2.7 years follow-up, individuals with diffuse malignant Parkinson's disease had greater progression in global composite outcome, motor symptoms, activities of daily living, and several non-motor features such as somnolence, depression, anxiety and cognition (including a decline of MoCA differing by almost two units). On the opposite side of the spectrum, over 50% of patients were classified as the mild motor-predominant

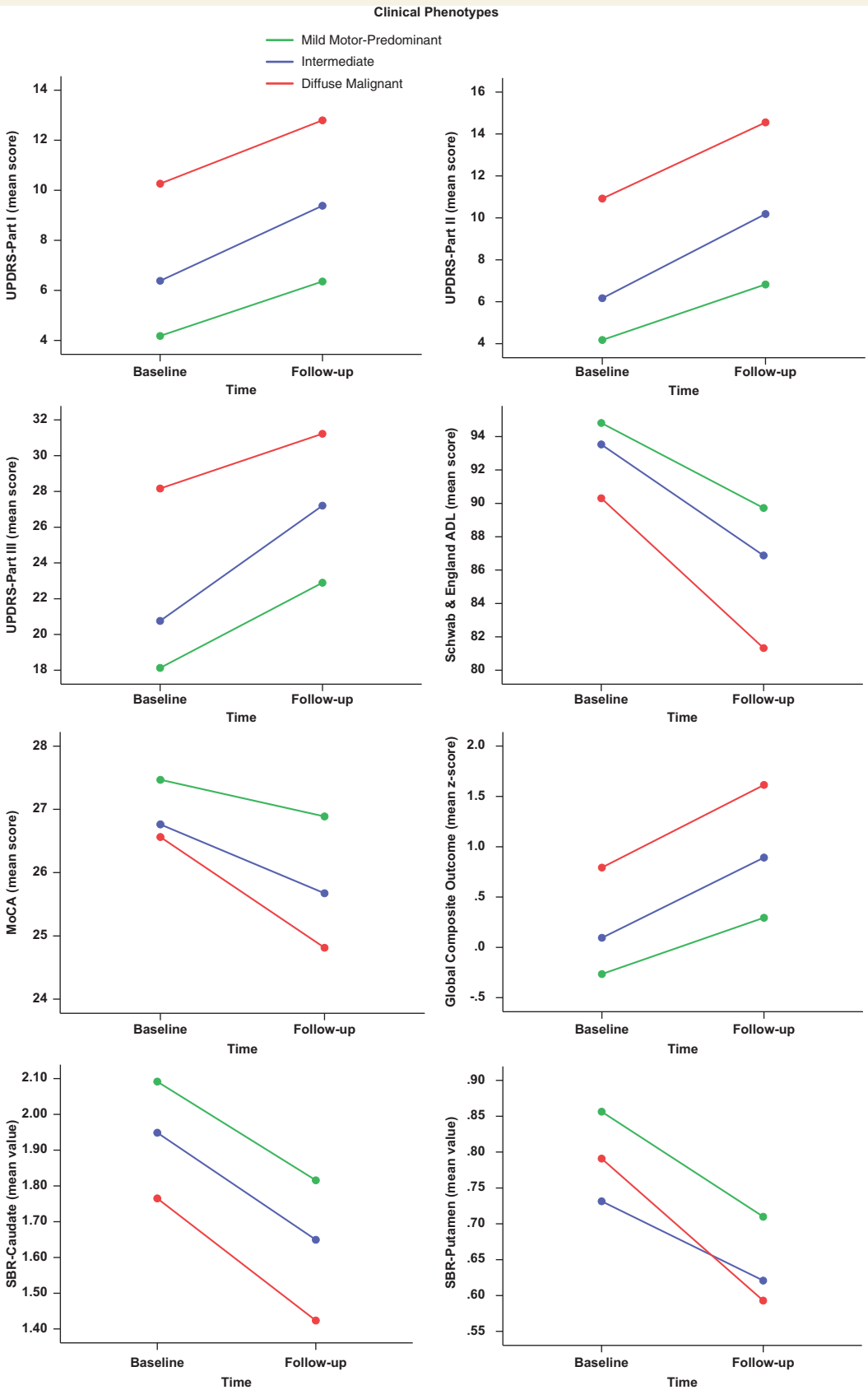


**Figure 4 Longitudinal changes in outcomes of interest in different phenotypes of Parkinson's disease in the PPMI population with at least 1 year of follow-up.** Mean follow-up duration in the entire population = 2.7 years. SBR = striatal binding ratio.

subtype. These had relatively modest impairment of motor and all major non-motor features at baseline. These patients had slower progression; even after 2.7 years they did not reach the baseline severity of the diffuse malignant subtype. The middle group, labelled intermediate, consisted of 35% of the sample, and had both baseline severity and progression rate between the other two extremes. For comparing the slope of progression between the subtypes, we adjusted for baseline values to overcome regression towards mean bias (Vickers and Altman, 2001) and to account for non-linearity of the scales (e.g. the MDS-UPDRS is designed to be sensitive to change at low scores). This makes a considerable difference because the crude recorded values (Fig. 4), fail to show differences in the rate of progression in motor disabilities. Nevertheless, regardless of analysis technique, people with diffuse malignant Parkinson's disease more rapidly progressed to severe cognitive impairment, severe sleep disturbance and restricted ADL. It will be of considerable interest to continue following these subtypes, to observe to what degree these differences in progression persist over the longer term.

Finally, typical clustering solutions present cluster characteristics at the group level, using mean values, which make it impossible to place individuals into a distinct subgroup, and so apply the solutions to real-life practice. Of note, substantial disagreement was observed for membership in the statistical clusters versus the clinical subtypes, especially for the intermediate and diffuse malignant subtypes. This reflects the flawed performance of clustering results in assigning individuals to their subtypes. Indeed, *post hoc* comparisons and especially analysis of disease progression found that the categorical clinical definition for Parkinson's disease phenotyping better discriminated distinct subgroups of Parkinson's disease than did the simple cluster analysis. By recommending clear-cut criteria for translating information learned from the cluster analysis into clinical subtypes, our solution has the benefit of allowing Parkinson's disease subtyping in real-life practical application (Supplementary material, subtype calculator).

## Genetic, imaging and CSF biomarkers

Our study also included genetic information for clustering the Parkinson's disease population. Instead of entering individual gene information (which would be underpowered), we used a single indicator 'genetic risk score', which summarizes data on 30 Parkinson's disease-specific mutations. Use of a summary indicator also facilitates fair weighting of genetic markers relative to other 17 classifiers. In a model designed to distinguish Parkinson's disease from controls, the genetic risk score explained 13.6% of the variance (Nalls *et al.*, 2015). However, in contrast, our results showed that the role of currently known genetic markers, at least when combined, is much less influential than clinical features for determining subtype. Note that the number of individuals with *LRRK2* and *GBA* risk variants in PPMI was low, and so they could not be entered as independent variables in the clustering model; there may still exist important differences in the prognosis of patients with different genetic mutations.

In contrast to genetics, imaging biomarkers demonstrated clear differences between subtypes. DBM-MRI analysis showed differences in patterns of atrophy of a Parkinson's disease-specific network. Atrophy was greater in the diffuse malignant and less in the mild motor-predominant subtypes. The MRI analysis procedure here is based on a network-based degeneration model in Parkinson's disease (Zeighami *et al.*, 2015). It specifically assesses a Parkinson's disease-specific pattern of atrophy that occurs primarily in structures functionally connected to substantia nigra. To the best of our knowledge, this is the first time that MRI-based measures of atrophy were found to differentiate *de novo* Parkinson's disease subtypes. Results from our exploratory voxel-wise comparisons between the subtypes suggest a pattern of increasing atrophy from a mild motor-predominant to intermediate, and finally diffuse malignant subtype. Nevertheless, the exploratory analyses failed to remain statistically significant following conservative corrections for multiple comparisons; power may be limited by sample size (especially in the diffuse malignant subtype) and by the modest degree of atrophy seen in early stages of Parkinson's disease. In addition to MRI, the hierarchical rank of SPECT striatal binding ratio seen in these subtypes provides external validation of the clinical subtypes, and demonstrates the utility of SPECT imaging for assessing disease severity and Parkinson's disease subtypes.

CSF biomarkers also differed between the subtypes. Interestingly, individuals with diffuse malignant Parkinson's disease, who had the fastest cognitive decline, showed an Alzheimer's disease-like CSF profile with low amyloid-β and amyloid-β/t-tau ratio. This further justifies the importance of cognitive impairment in subtyping Parkinson's disease. Our results are also aligned with those of a recent study showing that PPMI participants with low cerebral amyloid-β had more impaired cognitive performance, reduced grey matter volume and higher frequency of *APOE* ε4 alleles (McMillan and Wolk, 2016), and also with studies in which even small amounts of brain amyloid on Pittsburgh compound-B PET scanning were associated with increased dementia risk in Parkinson's disease (Gomperts *et al.*, 2013).

## Comparison to other subtyping solutions

Several studies have attempted to divide Parkinson's disease into subtypes. Most early subtyping systems were based on classification variables that were selected *a priori*, including age, tremor versus akinetic-rigid predominance, or the relative prominence of motor versus non-motor features. Tremor-dominant motor phenotype for instance, is

generally considered to have a more favourable prognosis than the PIGD phenotype (Jankovic, 2008). However, early single-variable subtyping solutions have been challenged because of inconsistent reliability and confounding by disease stage (Berg *et al.*, 2014). In our study, most of those in the mild motor-predominant subtype, were categorized as tremor-dominant (79%). However, only one-third of the diffuse malignant subtype was PIGD dominant. Moreover, the tremor PIGD classification alone could not predict prognosis in this study, unlike our global subtyping solution.

Data-driven cluster analysis has the advantage of a hypothesis-free approach, allowing broader interrogation of classifiers and unbiased estimates of factor importance. The first cluster analysis was performed by Graham and Sagar (1999), who introduced three subtypes as 'motor only', 'motor and cognition' and 'rapid progression'. Several clustering solutions have identified two clusters of 'old age-at-onset and rapid disease progression' and 'young age-at-onset and slow disease progression' (van Rooden *et al.*, 2011). Of crucial importance, cluster analyses are only as good as the variables they include. Most previous studies had a dearth of data on several important features particularly many non-motor and biomarker profiles, a limitation overcome by the comprehensiveness of the PPMI cohort. Also, most previous studies lacked longitudinal follow-up to enable the validation that comes with comparison of disease progression between different Parkinson's disease subtypes over time. Recently, using a separate cohort, we recommended a clustering-derived subtyping where three critical non-motor features—mild cognitive impairment, RBD and orthostatic hypotension—at baseline identified the most rapidly progressive subtype (also named 'diffuse malignant') over 4.5 years (Fereshtehnejad *et al.*, 2015). Interestingly, the results from both cohort studies, though with different baseline disease duration and length of follow-up, are in broad agreement, further highlighting the importance of non-motor manifestations as drivers of Parkinson's disease subtyping and prognosis.

A recent study by Erro *et al.* (2016) reported a non-hierarchical cluster analysis also on the PPMI database. It identified three subgroups of Parkinson's disease where apathy and hallucinations were found to be the most important classifiers. We did not use apathy and hallucination as classifiers as these were measured only by a single MDS-UPDRS-Part I item with discrete scores (and since only 13 patients had hallucinations at baseline). Our *post hoc* analysis showed significant difference between subtypes for apathy, but not for hallucinations. The agreement in the membership of PPMI participants between the Erro clustering and our recommended clusters is relatively low (56%). Of note, their recommended clusters (Erro *et al.*, 2016) did not differ in several manifestations known to influence Parkinson's disease prognosis, such as dysautonomia, depression and cognition. Moreover, differences in motor impairment were relatively modest; for example, while the average baseline MDS-UPDRS-Part III was 24.9 in the most severely affected cluster in their analysis, the average 'diffuse malignant' score was 30.0 in our solution. This may be because the previous solution did not include important features that were key classifiers in our model, namely neuropsychological testing, somnolence, and orthostatic hypotension. Furthermore, instead of using MoCA as a simple global measure, we included data from the full neuropsychological examination, as the baseline MoCA had small variance with clear ceiling effects (rendering it too insensitive for cluster analysis). We have also included MRI measures and CSF biomarkers for *post hoc* comparisons, and have assessed longitudinal progression.

## Pathophysiological explanations

The most distinct subtype was the diffuse malignant with more severe motor and non-motor symptoms, more atrophy in substantia nigra-connected areas, more dopaminergic deficit on SPECT and reduced amyloid-β in CSF. This subtype is consistent with one identified in a post-mortem study in which 25% of Parkinson's disease cases had an early malignant, dementia-dominant course and severe neocortical degeneration (Halliday *et al.*, 2008). Of note, patients in this clinico-pathologic series also had a higher occurrence of amyloid pathology on autopsy. One explanation might be the existence of a synergistic interaction between synucleinopathy and tau and amyloid pathology (McMillan and Wolk, 2016). However, an additional Alzheimer-like process with amyloid accumulation need not be the only explanation. The key markers of the diffuse malignant subtype are related to the dysfunction of diverse anatomical areas and brain pathways, are highly inter-correlated (Postuma *et al.*, 2008, 2009; Kim *et al.*, 2012; Rolinski *et al.*, 2014) and are not typical clinical features of Alzheimer's disease [e.g. RBD is notably uncommon in Alzheimer's disease (Boeve *et al.*, 2013)]. There is also evidence that synuclein may spread differently in more malignant subtypes. For example, in a separate cohort study, the presence of RBD in Parkinson's disease was associated with a more diffuse and severe deposition of synuclein seen at autopsy (Postuma *et al.*, 2015).

In some staging models of Parkinson's disease such as that of Braak (Braak *et al.*, 2003), Parkinson's disease pathology gradually progresses from olfactory tracts/medulla through the brainstem and the higher cortical layers. Our MRI findings find a similar hierarchical pattern of deformation, with the diffuse malignant subtype demonstrating greater degeneration along the Parkinson's disease-specific network. As there was no difference in disease duration between the subtypes, the different subtypes are not simply different stages of disease. This suggests that spreading models although true in aggregate, still include patients with both different progression speed and different spread patterns. There may exist a spectrum between a relatively substantia nigra-predominant disease and a multi-pathway

diffuse neurodegenerative process, within which every individual with Parkinson's disease is located.

## Limitations and strengths

Some limitations of this study should be noted. Follow-up duration of PPMI is still relatively short (<3 years), so we could not assess prognosis of subtypes over longer periods. Generalizability might be limited by the characteristics of the PPMI cohort; for example, people with overt cognitive impairment or significant early postural instability were excluded. Furthermore, the PPMI population is generally highly educated and highly motivated (i.e. the demands of the study may select out those with apathy). These factors could potentially lead to underrepresentation of the diffuse malignant form compared to the general Parkinson's disease population. The lack of an objective assessment for some features (e.g. no polysomnographic documentation of RBD) may have also masked larger differences. Many imaging and CSF biomarkers had missing values, which may lead to underestimation of their importance in clustering. Like all such analyses, our recommended Parkinson's disease subtyping needs to be validated in other cohorts. Moreover, longer follow-up of the PPMI cohort will allow continued assessment of changes in brain-imaging and other biomarkers, as well as migration between subtypes.

On the other hand, the current study and particularly the PPMI database upon which it is based have several strengths. This is the largest study both in terms of sample size and comprehensiveness to explore Parkinson's disease subtypes. Given that PPMI recruited a drug-naïve early Parkinson's disease population from multiple sites, our findings should still be mostly generalizable to *de novo* Parkinson's disease. They also can be applied in real clinical practice; rather than only a statistical cluster report, we provided discrete clinical criteria to define different subtypes of Parkinson's disease. Finally, this is one of only a few studies to compare longitudinal trend of progression between different subtypes.

# Conclusion

In summary, based on an extensive analysis of PPMI, we introduce three clinical subtypes of Parkinson's disease: 'mild motor-predominant', 'intermediate' and 'diffuse malignant'. Patients with the diffuse malignant subtype have more prominent dopaminergic deficit, more atrophy in Parkinson's disease-specific brain networks, a more Alzheimer's disease-like CSF profile and progress more rapidly. Further exploration of the underlying pathophysiologic differences between various Parkinson's disease subtypes will shed light on the underlying mechanisms for this variability. Ultimately, this knowledge could be used to develop a more efficient personalized approach for clinical trials and treatment strategies for individuals with different subtypes of Parkinson's disease.

# Acknowledgement

The authors would like to thank Chun Yao, PhD student in our research team, for his generous help and efforts to prepare the Excel 'Subtype Calculator'.

# Funding

This work was funded by grants from the Michael J Fox Foundation for Parkinson's Research, the W Garfield Weston Foundation, and the Alzheimer's Association, the Canadian Institutes for Health Research, the Natural Sciences and Engineering Research Council of Canada and the Preston Robb Fellowship in Canada. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit: www.ppmi-info.org. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including including AbbVie, Avid Radiopharmaceuticals, Biogen, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline (GSK), Eli Lilly and Company, Lundbeck, Merck, Meso Scale Discovery (MSD), Pfizer, Piramal Imaging, Roche, Sanofi Genzyme, Servier, Teva, and UCB (www.ppmi-info.org/fundingpartners).

# Supplementary material

Supplementary material is available at *Brain* online.

# References

Benton AL, Varney NR, Hamsher KD. Visuospatial judgment. A clinical test. Arch Neurol 1978; 35: 364–7.

Berg D, Postuma RB, Bloem B, Chan P, Dubois B, Gasser T, et al. Time to redefine PD? Introductory statement of the MDS task force on the definition of Parkinson's disease. Mov Disord 2014; 29: 454–62.

Boeve BF, Silber MH, Ferman TJ, Lin SC, Benarroch EE, Schmeichel AM, et al. Clinicopathologic correlations in 172 cases of rapid eye movement sleep behavior disorder with or without a coexisting neurologic disorder. Sleep Med 2013; 14: 754–62.

Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003; 24: 197–211.

Doty RL, Shaman P, Kimmelman CP, Dann MS. University of Pennsylvania smell identification test: a rapid quantitative olfactory function test for the clinic. Laryngoscope 1984; 94 (2 Pt 1): 176–8.

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.

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.

Formann AK. Die latent-class-analyse: einfuührung in die theorie und anwendung. Weinheim: Beltz; 1984.

Gladsjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, Heaton RK. Norms for letter and category fluency: demographic corrections for age, education, and ethnicity. Assessment 1999; 6: 147–78.

Gomperts SN, Locascio JJ, Rentz D, Santarlasci A, Marquie M, Johnson KA, et al. Amyloid is linked to cognitive decline in patients with Parkinson disease without dementia. Neurology 2013; 80: 85–91.

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.

Halliday G, Hely M, Reid W, Morris J. The progression of pathology in longitudinally followed patients with Parkinson's disease. Acta Neuropathol 2008; 115: 409–15.

Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008; 79: 368–76.

Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991; 14: 540–5.

Kang JH, Irwin DJ, Chen-Plotkin AS, Siderowf A, Caspell C, Coffey CS, et al. Association of cerebrospinal fluid beta-amyloid 1–42, T-tau, P-tau181, and alpha-synuclein levels with clinical features of drug-naive patients with early Parkinson disease. JAMA Neurol 2013; 70: 1277–87.

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.

Kim JS, Oh YS, Lee KS, Kim YI, Yang DW, Goldstein DS. Association of cognitive dysfunction with neurocirculatory abnormalities in early Parkinson disease. Neurology 2012; 79: 1323–31.

McMillan CT, Wolk DA. Presence of cerebral amyloid modulates phenotype and pattern of neurodegeneration in early Parkinson's disease. J Neurol Neurosurg Psychiatry 2016; 87: 1112–22.

Nalls MA, McLean CY, Rick J, Eberly S, Hutten SJ, Gwinn K, et al. Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. Lancet Neurol 2015; 14: 1002–9.

Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005; 53: 695–9.

Parkinson Progression Marker Initiative. The Parkinson progression marker initiative (PPMI). Prog Neurobiol 2011; 95: 629–35.

Postuma RB, Adler CH, Dugger BN, Hentz JG, Shill HA, Driver-Dunckley E, et al. REM sleep behavior disorder and neuropathology in Parkinson's disease. Mov Disord 2015; 30: 1413–17.

Postuma RB, Gagnon JF, Vendette M, Charland K, Montplaisir J. Manifestations of Parkinson disease differ in association with REM sleep behavior disorder. Mov Disord 2008; 23: 1665–72.

Postuma RB, Gagnon JF, Vendette M, Montplaisir JY. Markers of neurodegeneration in idiopathic rapid eye movement sleep behaviour disorder and Parkinson's disease. Brain 2009; 132 (Pt 12): 3298–307.

Rolinski M, Szewczyk-Krolikowski K, Tomlinson PR, Nithi K, Talbot K, Ben-Shlomo Y, et al. REM sleep behaviour disorder is associated with worse quality of life and other non-motor features in early Parkinson's disease. J Neurol NeurosurgPsychiatry 2014; 85: 560–6.

Schrag A, Barone P, Brown RG, Leentjens AF, McDonald WM, Starkstein S, et al. Depression rating scales in Parkinson's disease: critique and recommendations. Mov Disord 2007; 22: 1077–92.

Shapiro AM, Benedict RH, Schretlen D, Brandt J. Construct and concurrent validity of the Hopkins verbal learning test-revised. Clin Neuropsychol 1999; 13: 348–58.

Sieber BA, Landis S, Koroshetz W, Bateman R, Siderowf A, Galpern WR, et al. Prioritized research recommendations from the National Institute of Neurological Disorders and Stroke Parkinson's disease 2014 conference. Ann Neurol 2014; 76: 469–72.

Smith A. Symbol digit modalities test. Los Angeles, CA: Western Psychological Services; 1991.

Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.

Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: comparison with the unified Parkinson's disease rating scale. Mov Disord 2013; 28: 668–70.

Stiasny-Kolster K, Mayer G, Schafer S, Moller JC, Heinzel-Gutenbrunner M, Oertel WH. The REM sleep behavior disorder screening questionnaire–a new diagnostic instrument. Mov Disord 2007; 22: 2386–93.

van Rooden SM, Colas F, Martinez-Martin P, Visser M, Verbaan D, Marinus J, et al. Clinical subtypes of Parkinson's disease. Mov Disord 2011; 26: 51–8.

van Rooden SM, 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.

Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. BMJ 2001; 323: 1123–4.

Visser M, Marinus J, Stiggelbout AM, Van Hilten JJ. Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. Mov Disord 2004; 19: 1306–12.

Wechsler D. Wechsler adult intelligence scale. . San Antonio, TX: Psychological Corporation; 2008.

Weintraub D, Hoops S, Shea JA, Lyons KE, Pahwa R, Driver-Dunckley ED, et al. Validation of the questionnaire for impulsive-compulsive disorders in Parkinson's disease. Mov Disord 2009; 24: 1461–7.

Weintraub D, Simuni T, Caspell-Garcia C, Coffey C, Lasch S, Siderowf A, et al. Cognitive performance and neuropsychiatric symptoms in early, untreated Parkinson's disease. Mov Disord 2015; 30: 919–27.

Yesavage JA, Sheikh JI. 9/Geriatric Depression Scale (GDS). Clin Gerontol 1986; 5: 165–73.

Zeighami Y, Ulla M, Iturria-Medina Y, Dadar M, Zhang Y, Larcher KM, et al. Network structure of brain atrophy in *de novo* Parkinson's disease. Elife 2015; 4.