**Parkinson’s disease subtypes identified from cluster analysis of motor and non-motor symptoms**

Jesse Mu1; Kallol Ray Chaudhuri, MD2; Concha Bielza, PhD3; Jesus de Pedro-Cuesta, MD, PhD4; Pedro Larrañaga, PhD3; Pablo Martinez-Martin, MD, PhD4

1Department of Computer Science, Boston College, Chestnut Hill, Massachusetts, USA

2National Parkinson Foundation International Centre of Excellence, King’s College Hospital and King’s College; [NIHR Maudsley Biomedical Research Centre (BRC](http://www.maudsleybrc.nihr.ac.uk/)), London, UK

3Department of Artificial Intelligence, Technical University of Madrid, Madrid, Spain

4National Center of Epidemiology and CIBERNED, Carlos III Institute of Health, Madrid, Spain

**Corresponding author**

**K Ray Chaudhuri**

The Maurice Wohl Clinical Neuroscience Institute  
King's College London  
Cutcombe Road  
London   
SE5 9RT  
E-mail: [ray.chaudhuri@nhs.net](mailto:ray.chaudhuri@nhs.net)

Phone: + 44 0203 299 7153/7154

**Running title**: Parkinson’s disease cluster analysis

# Abstract

Parkinson’s disease is now considered a complex, multi-peptide, central and peripheral nervous system disorder with considerable clinical heterogeneity. Non-motor symptoms play a key role in the trajectory of Parkinson’s disease, from prodromal premotor to end stages. To understand the clinical heterogeneity of Parkinson’s disease, this study used cluster analysis to search for subtypes from a large, multi-centre, international, and well-characterized cohort of Parkinson’s disease patients across all motor stages, using a combination of cardinal motor features (bradykinesia, rigidity, tremor, axial signs) and, for the first time, specific validated rater-based non-motor symptom scales. Two independent international cohort studies were used: (a) the validation study of the Non-Motor Symptoms Scale () (Martinez-Martin *et al*., 2009a) and (b) baseline data from the global Non-Motor International Longitudinal Study () (Ray Chaudhuri *et al*., 2013). *k*-means cluster analyses were performed on the non-motor and motor domains (domains clustering) and the 30 individual non-motor symptoms alone (symptoms clustering), and average-linkage hierarchical agglomerative clustering was performed to group symptoms together. Four clusters are identified from the domains clustering supporting previous studies: mild, non-motor dominant, motor-dominant, and severe. In addition, six new smaller clusters are identified from the symptoms clustering, each characterized by clinically relevant non-motor symptoms. The clusters identified in this study present statistical confirmation of the increasingly important role of non-motor symptoms in Parkinson’s disease heterogeneity.

*Keywords:* Parkinson’s disease, subtypes, non-motor symptoms, cluster analysis

# Introduction

Parkinson’s disease (PD) is classically considered a motor disorder, with resting tremor, rigidity, and bradykinesia as its core features. However, the concept of Parkinson’s disease has changed considerably in the last few years, now prompting a revision of its diagnostic criteria to include non-motor symptoms (NMS) in the core parameters (Marras and Chaudhuri 2016; Postuma *et al.* 2015). There has been growing recognition that NMS in PD are caused by multi neurotransmitter pathway dysfunction which involves not just the central but also the peripheral nervous system (Jellinger 2012; Gjerløff *et al*. 2015).

Previous cluster analyses have already identified several non-motor clusters in PD patients (e.g. Erro *et al.,* 2013; Pont-Sunyer *et al*., 2015), while clinical observational studies also suggest the existence of specific non-motor subtypes (Sauerbier *et al.,* 2016; Marras and Chaudhuri 2016).

It has been argued that the recent concept of non-motor endophenotypes of PD provides a stronger basis for subtyping, since these relate to the central pathophysiology of specific neurotransmitter systems and are therefore likely to remain stable over time (Marras and Chaudhuri 2016). As such, several studies have explored PD subtypes while considering motor subtypes and their association with non-motor facets of the disease such as psychopathology and cognition (Graham and Sagar, 1999; Reijnders *et al*., 2009; Selikhova *et al*., 2009; Burn *et al*., 2012; Flensborg Damholdt *et al*., 2012), REM sleep behavior disorder (Romenets *et al*., 2012); and visual daily activities (Seichepine *et al*. 2011).

The present study was aimed at using cluster analysis techniques to search for PD subtypes from a large, multi-centre, international, and well-characterized cohort of PD patients across all stages, using a combination of motor cardinal features (bradykinesia, rigidity, tremor, axial signs) and comprehensive NMS assessed using specific validated rater-based scales. We believe this is the first study with these characteristics.

# Materials and Methods

## *Design*

Data from two independent international studies were used in the analysis: the validation study of the Non-Motor Symptoms Scale (NMSS) () (Martinez-Martin *et al*., 2009a) and baseline data from the global Non-Motor International Longitudinal Study (NILS) () (Ray Chaudhuri *et al*., 2013). NILS is a global initiative of the MDS Non-motor Study Group, has been adopted as a national study by the National Institute of Health Research in the UK (UKCRN No: 10084), and is a 5-year follow up study addressing the range, nature, and natural history of NMS in PD across all motor stages. All data in NILS are anonymised and centre-based datasets are entered into a secure international database for central analysis at the National Center of Epidemiology, Carlos III Institute of Health (Madrid, Spain).

## *Patients*

PD patients diagnosed according to internationally recognized criteria (Gibb and Lees, 1988; Lees *et al*., 2009), across all disease stages were included. For the NMSS study, patients had to be older than 30 years, but for inclusion of NILS patients there was no age limit. Exclusion criteria were: inability to read, understand, or answer written questionnaires; comorbidity, sequelae, or any disorder interfering with the assessment of PD manifestations; and inability to give informed consent. Patients with diagnoses other than Parkinson’s disease, such as MSA, PSP or DLB, were excluded. Patient recruitment was carried out across 15 countries in America, Asia, and Europe from 2007 to 2011.

## *Assessments*

For all patients, socio-demographic and historical data were recorded and the following assessments were applied:

1. The Scales for Outcomes in Parkinson’s Disease-Motor (SCOPA-Motor). A PD-specific rating scale with three dimensions: A. Examination (10 items); B. Activities of daily living (ADL, 7 items); and C. Complications (4 items). Each item scores from 0 (normal) to 3 (severe), the total score running from 0 to 75. This scale was derived from the Unified Parkinson’s Disease Rating Scale and showed high correlation coefficients with the original scale () and satisfactory clinimetric attributes in validation studies (Marinus *et al*., 2004; Martinez-Martin *et al*., 2005).
2. The Non-Motor Symptoms Scale (NMSS), a 30-item scale with nine domains: cardiovascular (2 items), sleep/fatigue (4 items), mood/apathy (6 items), perceptual problems/hallucinations (3 items), attention/memory (3 items), gastrointestinal tract (3 items), urinary function (3 items), sexual function (2 items), and miscellaneous (4 items). Each item scores from 0 to 12 (severity, 0 to 3, multiplied by frequency, 1 to 4) and the total NMSS score varies from 0 to 360 (Chaudhuri *et al*., 2007; Martinez-Martin *et al*., 2009a), a value representing the total non-motor “symptomatic burden” (NMSB).
3. The original Hoehn and Yahr (HY) scale (Hoehn and Yahr, 1967).
4. The Clinical Impression of Severity Index for PD (CISI-PD), a global evaluation of motor signs, disability, motor complications, and cognitive status. Items are rated from 0, normal, to 6, very severe, the total score ranging from 0 to 24 (Martinez-Martin *et al*., 2006; Martinez-Martin *et al*., 2009b).

## *Ethical aspects*

The NMSS validation study received ethical approval from the Carlos III Institute of Health, Madrid, Spain and local research ethics committees (Martinez-Martin *et al*., 2009a). The NILS is included in the UK Department of Health portfolio of approved studies (UK CRN portfolio Nr. 10084) and has been approved at all relevant institutions and the respective local ECs/IRBs of the participant researchers. Patients signed their informed consent before inclusion.

## *Statistical analysis*

Descriptive statistics (central tendency and dispersion, percentages) were applied to understand the main characteristics of the data. SCOPA-Motor examination items were grouped to obtain four aggregate “cardinal motor signs”: tremor (sum of items 1 and 2), bradykinesia (item 3), rigidity (item 4), and axial signs (sum of items 5 to 10). Additionally, an aggregate “motor complications” variable was obtained from the sum of items 18 to 21.

All variables were standardized before clustering, and unstandardized afterwards for interpretation. Analyses were conducted in R version 3.2.4 (www.r-project.org) and Stata version 14 (http://www.stata.com/)**.**

## Cluster analysis

*k*-means was used for cluster analysis. We performed two analyses on the patients in the dataset: the first clustering on the nine aggregate non-motor symptom domains, the four cardinal motor signs (tremor, bradykinesia, rigidity, axial), and motor complications, henceforth the “domains clustering”, and the second on the 30 individual non-motor symptoms of the NMSS only, henceforth the “symptoms clustering”. Average-linkage hierarchical agglomerative clustering on the 30 non-motor symptoms, 4 motor signs, and motor complications was also performed to observe the grouping of the symptoms.

Various formal measures were used to determine the optimal number of clusters for the dataset. For the domains clustering, the optimal according to the Gap Statistic and the 1-standard-error method suggested by Tibshirani *et al*. (2001) was (Supplementary Fig. 1A). Other cluster determination methods (within sum of squares error scree plot, minimum average silhouette width) suggested where simply divided the data uninformatively into groups with varying levels of overall PD severity. Thus was selected to offer a good blend of model fit, informativeness, and parsimony. The same method was applied for the symptoms clustering, and the optimal number of clusters indicated was (Supplementary Fig. 1B). was tried to compare results to the domains clustering, but results were not interpretable.

## Comparative subgroup analysis

For both clusterings, we displayed symptoms with a heatmap, scaling cluster averages to z-scores for each variable. Additionally, for the domains clustering, we displayed the four clusters using boxplots, which allowed us to visualize the center and spread of symptoms for each cluster. Finally, for the hierarchical clustering on the symptoms themselves, we displayed results in a dendrogram and clustered the symptoms into six interpretable clusters.

For each variable in both clusterings, we used one-way ANOVA and tests to respectively check the equality of variable means and proportions across the clusters found, using Bonferroni correction for multiple testing with corrected *p*-value considered significant. Differences among pairwise clusters were tested post-hoc using Tukey’s range test for continuous means, or pairwise tests for proportions, with Bonferroni correction both for the within-variable pairwise tests and the multiple variable comparisons.

To compare the domains and symptoms clusterings, we depicted cluster alignment with a contingency table, and computed the adjusted rand index (ARI; Hubert and Arabie, 1985) and *v*-measure (Rosenberg and Hirschberg, 2007), to evaluate similarity between the two clusterings. Both measures range from 0 (no similarity) to 1 (identical), and respectively take a contingency table-based and information-theoretic approach to measuring clustering similarity.

Lastly, to explore the relationship between domain and symptom severity and disease duration, we computed the correlation coefficient for each variable with PD duration and fitted smoothed loess curves to the data both globally and for each cluster found in the domains clustering.

# Results

***Study sample***

Out of the 951 patients in the study, we used listwise deletion to exclude 47 patients due to missing measurements in the variables of interest, resulting in 904 remaining patients. There were no significant differences between the included and excluded groups with respect to age, sex, PD duration, and HY (). The characteristics of the sample included for analysis () are displayed in Table 1. Patients were predominantly male (62.17%). 13.38% were in HY stage 1; 43.36% in stage 2; 29.65% in stage 3; 11.50% in stage 4; and 2.10% in stage 5.

Table 1. Description of the sample.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mean | SD | Median | Range |
| Age at study | 64.28 | 9.86 | 65 | 34 – 89 |
| Age at onset of Parkinson’s disease (PD onset) | 56.27 | 10.72 | 57 | 25 – 89 |
| Duration of the disease (PD duration) | 8.01 | 5.80 | 7 | 0 – 40 |
| Non-Motor Symptoms Scale total score | 50.45 | 41.72 | 39 | 0 – 225 |
| Cardiovascular  Sleep/Fatigue  Mood/Apathy  Perceptual problems/Hallucinations  Attention/Memory  Gastrointestinal  Urinary  Sexual function  Miscellaneous | 1.74  8.76  8.67  1.64  5.40  5.53  8.07  3.53  7.12 | 3.26  8.71  11.54  3.86  7.42  6.78  8.93  5.98  7.78 | 0  6  4  0  2  3  5  0  4 | 0 – 24  0 – 48  0 – 60  0 – 33  0 – 36  0 – 36  0 – 36  0 – 24  0 – 48 |
| Cardinal motor features \*  Tremor  Bradykinesia  Rigidity  Axial | 2.59  2.40  2.23  3.25 | 2.58  1.41  1.36  2.67 | 2  2  2  3 | 0 – 12  0 – 6  0 – 6  0 – 12 |
| SCOPA-Motor Total score | 21.07 | 12.06 | 19 | 1 – 72 |
| A. Examination  B. Activities of daily living  C. Complications | 11.54  6.80  2.73 | 6.56  4.19  3.01 | 10  7  2 | 0 – 41  0 – 21  0 – 12 |
| ­­­­Clinical Impression of Severity Index | 8.29 | 4.61 | 8 | 0 – 24 |
| SD: Standard deviation. SCOPA: Scales for Outcomes in Parkinson’s Disease.  \* Scores derived from items of the SCOPA-Motor A. Examination. | | | | |

## *Domains clustering*

Results from the *k*-means clustering on the nine non-motor domains, the four cardinal motor signs, the motor complications, as well as additional variables not used in the analysis, are reported in Table 2 (heatmap in Fig.1; boxplots in Supplementary Fig. 2). Cluster means for all variables were found to be statistically different except for age at PD onset and sex ( after correcting for the comparisons of the 19 variables). Specific pairwise differences are noted in the table.

Cluster 1 (D1, ) patients were mildly affected in all domains. This cluster was characterized by relatively lower disease durations and ages.

Cluster 2 (D2, ) patients were severely affected in non-motor domains but mildly affected in motor domains. This cluster had a severity of motor variables relatively similar to the cluster 1 (mild) subtype especially in tremor, but generally expressed significantly higher scores for non-motor domains than clusters 1 and 3, especially in the sleep/fatigue, mood/apathy, urinary, and miscellaneous domains. In motor complications uniquely, scores were not statistically different from cluster 3.

Cluster 3 (D3, ) patients were severely affected in motor domains but mildly affected in non-motor domains. Mean motor scores were greater than the means of clusters 1 and 2, with the exception of motor complications. Additionally, mean motor scores were less than 4, with the exception of tremor, which was especially high. Importantly, CISI-PD scores of clusters 2 and 3 were not statistically significantly different, and no differences were observed in cluster 2 and cluster 3 age or PD duration.

Cluster 4 (D4, ) patients were severely affected in all domains, having the greatest symptom mean out of all four clusters with the exception of tremor. Consequently, patients in Cluster 4 had the longest average PD duration and older ages, but did not have a significantly different age of PD onset.

Figure 1. **Domains clustering heatmap**. Heatmap of symptoms for each cluster, separated by white lines according to 9 non-motor domains, 4 cardinal motor features, motor complications, and 4 variables not included in the analyses. Since symptoms have different scales, cluster means for each symptom are displayed as standardized scores relative to each overall symptom mean.

Table 2. Domains clustering summary. Unless otherwise specified, statistics are reported as mean (sd).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Cluster | 1 | 2 | 3 | 4 |
|  | *n*(%) | 428 (47%) | 180 (20%) | 232 (26%) | 64 (7%) |
| Non-motor | 1. Cardiovascular | 0.7 (1.5)2,4 | 3.2 (3.7)1,3,4 | 1.1 (2.1)2,4 | 6.9 (6.4)1,2,3 |
| domains | 2. Sleep/fatigue | 4.5(5.0)2,3,4 | 16 (8.7)1,3,4 | 7.5 (6.6)1,2,4 | 21.7 (9.7)1,2,3 |
|  | 3. Mood/apathy | 3.4 (4.8)2,3,4 | 19.2 (15.0)1,3 | 6.6 (8.0)1,2,4 | 21.7 (13.5)1,3 |
|  | 4. Perception/hallucination | 0.5 (1.7)2,4 | 2.7 (4.3)1,3,4 | 0.8 (1.8)2,4 | 9.7 (6.9)1,2,3 |
|  | 5. Attention/memory | 3.0 (4.5)2,4 | 10.5 (9.2)1,3,4 | 3.3 (4.4)2,4 | 14.5 (11.0)1,2,3 |
|  | 6. Gastrointestinal | 2.9 (4.1)2,3,4 | 8.5 (7.1)1,3,4 | 4.7 (5.3)1,2,4 | 17.4 (9.2)1,2,3 |
|  | 7. Urinary | 4.7 (6.2)2,4 | 14.0 (9.9)1,3,4 | 6.2 (6.7)2,4 | 20.3 (9.7)1,2,3 |
|  | 8. Sexual function | 1.7 (3.4)2,3,4 | 7.3 (7.8)1,3 | 2.4 (4.1)1,2,4 | 9.0 (9.9)1,3 |
|  | 9. Miscellaneous | 4.0 (4.8)2,3,4 | 13.2 (8.7)1,3 | 6.2 (6.8)1,2,4 | 14.5 (10.1)1,3 |
| Motor symptoms | Axial | 1.7 (1.5)2,3,4 | 3.6 (2.2)1,3,4 | 4.5 (2.3)1,2,4 | 8.2 (2.7)1,2,3 |
|  | Bradykinesia | 1.6 (0.9)2,3,4 | 2.2 (1.1)1,3,4 | 3.5 (1.0)1,2,4 | 4.5 (1.3)1,2,3 |
|  | Rigidity | 1.5 (0.9)2,3,4 | 1.8 (1.2)1,3,4 | 3.3 (1.0)1,2,4 | 4.2(1.2)1,2,3 |
|  | Tremor | 2.0 (1.9)3,4 | 1.5 (1.8)3,4 | 4.1 (2.8)1,2 | 4.1 (4.1)1,2 |
|  | Motor complications | 1.4 (2.1)2,3,4 | 3.1 (2.9)1,4 | 3.7 (2.9)1,4 | 7.0 (3.6)1,2,3 |
| Variables not | Sex (% male) | 64 | 54 | 67 | 58 |
| used in analysis | CISI-PD total | 5.5 (3.0)2,3,4 | 9.6 (3.8)1,4 | 10.1 (3.5)1,4 | 16.4 (4.6)1,2,3 |
|  | Age | 62.5 (9.7)4 | 65.2 (9.4)4 | 64.9 (10.1)4 | 71.1 (7.9)1,2,3 |
|  | PD onset | 56 (10.5) | 56.6 (10.6) | 56.3 (11.3) | 56.7 (10.6) |
|  | PD duration | 6.5 (4.7)2,3,4 | 8.6 (5.7)1,4 | 8.6 (5.7)1,4 | 14.4 (8.0)1,2,3 |
| 1 Significant difference with cluster D1 (corrected )  2 Significant difference with cluster D2 (corrected )  3 Significant difference with cluster D3 (corrected )  4 Significant difference with cluster D4 (corrected ) | | | | | |

## *Symptoms clustering*

*k*-means performed on the 30 individual non-motor symptoms found 6 clusters, which were ordered according to increasing CISI-PD score (Table 3, heatmap in Fig. 2). Means of all symptoms were found to differ across clusters except for PD onset, sex, and tremor, with pairwise differences again noted in the table.



Figure 2. **Symptoms clustering heatmap**. Heatmap of symptoms for each cluster, separated by white lines according to 30 individual non-motor symptoms, 4 cardinal motor features, motor complications, and 4 variables not included in the analysis. Since symptoms have different scales, cluster means for each symptom are displayed as standardized scores relative to each overall symptom mean.

Cluster 1 (S1, ), the largest cluster representing 50% of the group, was similar to domains cluster 1, and was composed of patients relatively mildly affected in all NMS. Cluster 2 (S2, ) had higher mean symptom scores than Cluster 1’s in several cases, including restless legs syndrome (RLS), swallowing, and the miscellaneous domain, but could nonetheless be classified as a mild/moderate cluster.

Although clusters 3–6 increased in motor and overall PD severity, they varied significantly in their non-motor expression and generally expressed a specific set of non-motor symptoms unique to the cluster. These characterizing groups of non-motor symptoms aligned well with the established non-motor domains. Cluster 3 (S3, ) mainly expressed domain 7 (urinary), while cluster 4 (S4, ) was affected severely in domain 3 (mood/apathy). Cluster 5 (S5, ) showed severe impact in most non-motor symptoms but especially in domain 5 (attention/memory). Similarly, cluster 6 (S6, ) revealed severe effects across all non-motor symptoms and motor features, but was most severely affected in cardiovascular, perception/hallucination, and gastrointestinal NMSS domains. Overall, the symptoms clustering appeared to fragment several of the domains clusters into smaller groups, as suggested in the next section.

Table 3. Symptoms clustering summary. Unless otherwise specified, statistics are reported as mean(sd).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cluster | 1 | 2 | 3 | 4 | 5 | 6 |
|  | *n*(%) | 456 (50%) | 201 (22%) | 100 (11%) | 73 (8%) | 54 (6%) | 20 (2%) |
| 1. | Lightheadedness | 0.5 (1.1)2,3,4,5,6 | 1.7 (2.5)1,5,6 | 2.5 (3.4)1,6 | 1.9 (2.5)1,6 | 3.3 (3.7)1,2,6 | 7.7 (3.6)1,2,3,4,5 |
| Cardiovascular | Fainting | 0.1 (0.6)6 | 0.3 (1)6 | 0.1 (0.5)6 | 0.5 (1.2)6 | 0.1 (0.5)6 | 6.3 (2.9)1,2,3,4,5 |
| 2. Sleep/ | Drowsiness | 0.9 (1.8)2,3,4,5,6 | 2.4 (2.9)1,5,6 | 2.3 (2.9)1,5,6 | 2.2 (3)1,5,6 | 6.6 (4.2)1,2,3,4 | 5.2 (3)1,2,3,4 |
| fatigue | Fatigue | 1.2 (1.9)2,3,4,5,6 | 4.2 (3.7)1,4,5 | 4.5 (4)1,5 | 5.9 (4.2)1,2 | 8 (3.5)1,2,3 | 6.9 (3.1)1 |
|  | Insomnia | 1.1 (2.2)2,3,4,5,6 | 3 (3.8)1,4 | 3.1 (4.2)1,4 | 5.3 (4.8)1,2,3 | 4.7 (4.8)1 | 5 (2.4)1 |
|  | RLS | 0.5 (1.3)2,4,5,6 | 2.4 (3.5)1,3,6 | 1 (2.4)2,5,6 | 1.8 (3.2)1,6 | 3.2 (4.2)1,3 | 4.9 (2.6)1,2,3,4 |
| 3. Mood/ | Loss interest | 0.4 (0.9)2,4,5,6 | 1.1 (1.8)1,4,5,6 | 0.6 (1.3)4,5,6 | 6.6 (3.9)1,2,3,5 | 4.5 (3.7)1,2,3,4 | 5.3 (2.8)1,2,3 |
| apathy | Loss activities | 0.6 (1.3)2,4,5,6 | 1.9 (2.7)1,4,5,6 | 1 (2)4,5,6 | 7.8 (3.5)1,2,3,5,6 | 6 (4.5)1,2,3,4 | 4.7 (2.9)1,2,3,4 |
|  | Anxiety | 0.8 (1.6)2,4,5,6 | 2.2 (2.9)1,4,5,6 | 1.6 (2.9)4,5,6 | 5.8 (4.3)1,2,3 | 6.4 (4.4)1,2,3 | 4.7 (3)1,2,3 |
|  | Depression | 0.7 (1.4)2,4,5,6 | 2.7 (3.2)1,3,4,5,6 | 1.4 (2.3)2,4,5,6 | 7.4 (4)1,2,3 | 5.8 (4.3)1,2,3 | 5.3 (3.2)1,2,3 |
|  | Flat affect | 0.3 (1)2,4,5,6 | 1.1 (2.1)1,4,5 | 0.8 (2.1)4,5,6 | 4.8 (4.3)1,2,3,5 | 2.9 (3.7)1,2,3,4 | 3 (2.1)1,3 |
|  | Loss pleasure | 0.3 (1.1)4,5,6 | 1 (1.8)4,5,6 | 0.8 (1.8)4,5,6 | 7.3 (3.8)1,2,3,5,6 | 4.6 (4.2)1,2,3,4 | 3.6 (2.8)1,2,3,4 |
| 4. Perception/ | Hallucination | 0.2 (0.9)5,6 | 0.6 (1.7)5,6 | 0.7 (1.9)5,6 | 0.6 (1.8)5,6 | 2.7 (3.3)1,2,3,4,6 | 4.8 (3.3)1,2,3,4,5 |
| hallucination | Delusion | 0.1 (0.6)4,5,6 | 0.3 (1.4)4,5,6 | 0.1 (0.6)4,5,6 | 1.3 (3)1,2,3,6 | 2 (3.5)1,2,3,6 | 4.7 (3.4)1,2,3,4,5 |
|  | Diplopia | 0.2 (0.8)5,6 | 0.6 (1.4)5,6 | 0.4 (1.6)5,6 | 0.8 (2.2)5,6 | 3.1 (4.4)1,2,3,4 | 3.6 (2.7)1,2,3,4 |
| 5. Attention/ | Loss concentration | 0.7 (1.5)2,3,4,5,6 | 2.6 (2.9)1,5,6 | 2.2 (3.4)1,5,6 | 2.9 (2.9)1,5 | 7.5 (3.9)1,2,3,4 | 5.1 (2.8)1,2,3 |
| memory | Forget explicit | 0.7 (1.3)2,3,4,5,6 | 2.2 (2.7)1,5,6 | 2.2 (3)1,5,6 | 2.1 (2.5)1,5,6 | 8.5 (3.2)1,2,3,4,6 | 4.8 (2.9)1,2,3,4,5 |
|  | Forget implicit | 0.5 (1.2)2,3,4,5,6 | 1.9 (2.6)1,5,6 | 1.5 (2.7)1,5,6 | 1.8 (2.2)1,5,6 | 7.6 (4.1)1,2,3,4,6 | 5.1 (3.2)1,2,3,4,5 |
| 6. | Drooling | 0.6 (1.5)2,3,4,5,6 | 2.3 (3.3)1,5,6 | 3.3 (4.1)1,6 | 3.3 (4.2)1,6 | 4.2 (4.8)1,2 | 6.2 (3.3)1,2,3,4 |
| Gastrointestinal | Swallowing | 0.3 (0.8)2,3,5,6 | 2 (3)1,6 | 1.2 (2)1,6 | 1.2 (2)6 | 2.3 (2.9)1,6 | 6.5 (4.2)1,2,3,4,5 |
|  | Constipation | 1.5 (2.7)2,3,4,5,6 | 3 (3.8)1,6 | 3.8 (4.4)1,6 | 3.7 (4.4)1,6 | 4.5 (4.9)1,6 | 8.8 (4.4)1,2,3,4,5 |
| 7. Urinary | Urinary urgency | 0.9 (1.7)2,3,4,5,6 | 1.8 (2.4)1,3,5,6 | 6.6 (3.8)1,2,4 | 2.4 (3.4)1,3,5,6 | 7.7 (4.3)1,2,4 | 6.2 (3.5)1,2,4 |
|  | Urinary frequency | 0.9 (1.8)3,4,5,6 | 1.3 (1.9)3,4,5,6 | 7.7 (3.5)1,2,4 | 3.1 (3.6)1,2,3,5,6 | 6.2 (4.5)1,2,4 | 6.7 (3.1)1,2,4 |
|  | Nocturia | 1.7 (2.3)3,4,5,6 | 2.6 (2.9)3,4,5,6 | 8.5 (3.5)1,2,4 | 4.6 (4.2)1,2,3,5 | 6.9 (4.4)1,2,4 | 7.1 (3.4)1,2 |
| 8. Sexual | Sex drive | 0.7 (1.7)2,3,4,5,6 | 1.9 (3.4)1,5 | 3 (4.1)1,5 | 3.6 (4.4)1,5 | 6 (5.3)1,2,3,4 | 3.9 (5.3)1 |
|  | Sex dysfunction | 0.7 (2)3,4,5,6 | 1.8 (3.3)3,5 | 3.4 (4.3)1,2 | 2.3 (4.2)1,5 | 5 (5.2)1,2,4 | 3.8 (5.6)1 |
| 9. | Unexplained pain | 0.7 (1.8)2,3,4,5,6 | 2.6 (3.9)1 | 2.4 (4)1 | 2.3 (3.6)1 | 4.3 (5)1 | 4.3 (2.2)1 |
| Miscellaneous | Gustation/olfaction | 1.2 (2.5)2,4,5,6 | 4 (4.2)1 | 2.5 (3.6) | 3 (3.9)1 | 4 (4.7)1 | 5.5 (4.7)1 |
|  | Weight change | 0.8 (1.4)2,4,5,6 | 1.9 (3.1)1,5 | 1.8 (3)5 | 2.1 (2.8)1 | 3.6 (4.3)1,2,3 | 4 (3.9)1 |
|  | Sweating | 0.6 (1.6)2,3,4,5,6 | 2.5 (3.9)1 | 2.4 (3.9)1 | 2.1 (3.6)1 | 2.8 (3.9)1 | 3.9 (3.4)1 |
| Motor | Axial | 2.3 (2)2,3,4,5,6 | 3.5 (2.6)1,5,6 | 4 (2.8)1,6 | 4.4 (2.7)1,6 | 5.3 (3.3)1,2,6 | 8.5 (2.3)1,2,3,4,5 |
| symptoms | Bradykinesia | 2.1 (1.2)4,5,6 | 2.5 (1.4)6 | 2.5 (1.4)6 | 2.9 (1.4)1,6 | 3 (1.6)1,6 | 4.5 (1.7)1,2,3,4,5 |
|  | Rigidity | 2 (1.2)6 | 2.3 (1.5)6 | 2.3 (1.4)6 | 2.5 (1.3)6 | 2.6 (1.5)6 | 4.7 (1)1,2,3,4,5 |
|  | Tremor | 2.6 (2.4)6 | 2.5 (2.4)6 | 2.1 (2.5)6 | 2.5 (2.6)6 | 2.4 (2.8)6 | 5.8 (5.3)1,2,3,4,5 |
|  | Motor comp | 1.9 (2.5)2,4,5,6 | 3.5 (3)1,6 | 2.7 (2.7)6 | 3.7 (3.3)1,6 | 3.9 (3.7)1,6 | 8 (3.7)1,2,3,4,5 |
| Variables not | Sex | 67 | 57 | 64 | 49 | 63 | 50 |
| used in analysis | CISI PD total | 6.4 (3.5)2,3,4,5,6 | 9.3 (4.1)1,5,6 | 9.4 (4.1)1,6 | 10.7 (4.6)1,6 | 11.8 (5.5)1,2,6 | 18.4 (4.2)1,2,3,4,5 |
|  | Age | 62.8 (10)3,5,6 | 63.8 (10)5 | 68.3 (7.5)1 | 63.1 (9.2) | 69.7 (9.2)1,2 | 72.1 (7.7)1 |
|  | PD onset | 56.2 (10.8) | 55 (10.7) | 58.2 (10) | 54.8 (10.3) | 58.8 (11.7) | 59.2 (8.6) |
|  | PD duration | 6.6 (4.8)2,3,5,6 | 8.8 (5.9)1 | 10.1 (6.7)1 | 8.3 (4.9) | 11 (8.5)1 | 12.8 (6.5)1 |
| *n* Significant difference with cluster S*n* (corrected ) | | | | | | | |

## *Comparison between clusterings*

Alignment of the D and S clusters is visualized in Table 4. While S1 grouped patients from D1 (mild) and D3 (motor-dominant), and D4 (severe) showed a predominant contribution from S5 (severe non-motor) and S6 (severe motor and non-motor), the remaining clusters were fragmentarily distributed, as indicated by low similarity metrics (). For the domains clustering, patients in clusters mildly affected in non-motor domains (D1, D3) were correspondingly distributed among the milder symptoms clusters (S1–S4, skewed left). Conversely, patients in clusters with severe non-motor symptoms (D2, D4) were split among the various specific non-motor-dominant clusters (S2–S6), suggesting that the symptoms clustering is clinically more precise than the domains clustering. This is particularly evident from D2, where 95 out of the 180 patients were clustered based on urinary (S3) and mood/apathy (S4) symptoms.

Table 4. Contingency table describing cross-categorization of individuals in the domains and symptoms clusters.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Symptoms clusters | | | | | | |  |
| Domains clusters |  | S1 | S2 | S3 | S4 | S5 | S6 | Total |
| D1 | 335 | 64 | 26 | 3 | 0 | 0 | 428 |
|  | 0 | 54 | 46 | 49 | 31 | 0 | 180 |
|  | 121 | 77 | 22 | 12 | 0 | 0 | 232 |
| D4 | 0 | 6 | 6 | 9 | 23 | 20 | 64 |
| Total | | 456 | 201 | 100 | 73 | 54 | 20 |  |

***Hierarchical clustering on variables***

**Hierarchical clustering on the 30 non-motor symptoms and the four cardinal motor signs produced the dendrogram in Fig. 3. As expected, symptoms belonging to the same domain of the NMSS tended to cluster together, with some notable exceptions. Diplopia (domain 4) was grouped away from other perception/hallucination symptoms, instead categorized loosely with domain 8 (sexual). RLS (domain 2) was grouped with domain 9 (miscellaneous) symptoms, albeit at a very high point on the tree, indicating high dissimilarity. Similarly, drowsiness (domain 2) was grouped with the attention/memory symptoms of domain 5. Notably, tremor was the most isolated symptom, occupying a single branch at the top of the tree.

Figure 3. **Hierarchical clustering on symptoms.** Average-linkage hierarchical clustering of motor (blue) and non-motor (black) symptoms. Symptoms are labeled with their name and corresponding domain number. The dendrogram is colored with 5 clusters.

## *Correlation analysis*

Most variables had little to no correlation with PD duration (Supplementary Fig. 3) due to their high variance. Variables with the highest correlation, near 0.5, included urinary frequency, swallowing, and drooling; domains 9 (miscellaneous), 7 (urinary), and 6 (gastrointestinal); axial and motor complications; and total CISI-PD score. Scatterplots for CISI Total, Tremor, Anxiety, and Depression appear in Supplementary Fig. 4.

# Discussion

We believe that this is the first cluster analysis-based study of PD-related motor and non-motor symptoms from a large international multicenter cohort. Previous cluster analysis-based studies have either focused on early/untreated PD (Erro *et al*., 2013; Pont-Sunyer *et al*., 2015) or lack detailed assessments based on the severity and frequency of non-motor domains and symptoms (van Rooden *et al*., 2011).

The domains clustering produced 4 clusters (Fig. 1). Cluster D1 is a mild disease subtype with reduced disease duration, and is similar to clusters reported by Erro *et al.* (2013) and van Rooden *et al*. (2011). Cluster D2 corresponds to the non-motor dominant phenotype described in clinical phenotype driven studies (Sauerbier *et al*., 2016). Cluster D3 corresponds to the traditional motor-dominant view of Parkinson’s disease. However, the high incidence of tremor in D3 (higher than D4) is of interest and appears to reflect not only the motor-dominant subtype of van Rooden *et al*. (2011) but also the tremor-dominant/slow-progression as described in the study by Ma *et al*. (2015). Cluster D4 is a severe motor and non-motor disease subtype which replicates several previous studies.

From the symptoms clustering (Fig. 2), six smaller clusters were identified. S1 was similar to D1. Interestingly, S2 to S6, while increasing in motor severity, expressed specific NMS, thus supporting the clinical concept of NMS-based subtyping of PD. Cluster S2, with principal components including RLS, swallowing, pain, and others, may be a new finding from this study. Cluster S3, with high urinary dysfunction, fits the descriptions by Erro *et al*. (2013, 2015) highlighting the relevance of this symptom as a specific marker in the non-motor dominant cluster and also for disease progression. Cluster S4, characterized by high mood/apathy-based symptoms, is consistent with the sleep and apathy clinical phenotypes described by other studies (Sauerbier *et al*., 2016). Clusters S5 and S6 are of clinical interest, as in these clusters NMS are dominant, overshadowing motor symptoms with an emphasis on cognitive impairment in S5 and autonomic (cardiovascular and gastrointestinal) symptoms in S6.

The comparison between the domains and symptoms clustering shown in the contingency table (Table 4) suggests that the broader subset of S1, a mild non-motor dominant cluster, essentially expresses two NMS subtypes, one of them with motor symptoms. The low numbers observed in some cells do not allow consistent clinical interpretation. The dendrogram (Fig. 3) indicates that the symptoms grouping in the NMSS dimensions works as expected, as most items in each domain segregate together.

Our data mirror the observations of van Rooden *et al.*, (2011) who used different rating scales and modeling methods (probabilistic clustering based on the Expectation-Maximization algorithm). Unlike van Rooden, our data show that differences in mean disease durations do exist between mild (D1, S1) and severe (D4, S6) subtypes, likely due to advanced disease duration. However, S2 and S3 are newly reported and their characteristics support clinical endophenotyping of non-motor subtypes not reported in the van Rooden study.

The differences in disease progression for each cluster can be seen by the corresponding graphs in Supplementary Fig. 4. Non-motor dominant cluster D2 starts at higher scores for anxiety and depression, and actually decreases with increasing PD duration, thus indicating that these patients’ subtypes can be determined early after PD onset from the depressive symptom score. Similar improvements in NMS and depression and anxiety scores have been reported recently in a 2-year follow-up study of an untreated PD cohort (Mollenhauer *et al*., 2016).

What could be the clinical implications of these clusterings? Firstly, our analysis represents statistical conformation of the growing recognition of NMS-dominant presentation of PD and its heterogeneity. The specific expression of several NMS domains such as mood/anxiety, sleep/fatigue, cognition, and urinary function suggests that these subgroups may have different patterns of neurodegeneration involving the brain’s various non-dopaminergic pathways, possibly in excess of dopaminergic degeneration, as suggested by many authors (Jellinger 2012). Secondly, clinical recognition of subtypes using ad hoc criteria would allow for the development of truly subtype-specific treatment packages for PD (Marras and Chaudhuri 2016). Thirdly, clinical characterisation of these groups will allow studies of natural history of specific subtypes.

Due to the data collection methods of the two studies used in this analysis, selection due to prevalence bias, i.e. overrepresentation at the study sample of patients with higher survival, would not likely explain this clustering. In contrast, clustering at clinical early stages might have been undermined by neglect of symptoms poorly recorded prior to diagnosis. The clinical non-specificity of D1, S1, D4, and S6, with extremely different disease duration and severity, contrasts with the identification of clinically characterized S3, S4, and S5, clusters with dominant expression of urinary, mood/apathy and attention/memory symptoms, respectively, at intermediate stages of the disease course. This pattern is in line with the notion of “phenotypic convergence” proposed by Warren *et al*. (2013) as a key clinical feature of the spread of neurodegenerative disorders due to abnormal protein aggregates. The identified clusters may represent distinct footprints of large-scale network disintegration whose translation to clinical management is required.

Like any cohort-based, cluster-analysis driven study, there are several limitations of this analysis. First, we did not report a control group, although our intention was not to describe the symptoms as discriminant from normal subjects. Thus, the importance of including a control group could be questioned. Second, NMS profiles in advanced PD could also be influenced by patterns of dopaminergic therapy, particularly hallucinations and orthostatic problems. However, these symptoms did not emerge as key drivers of any of the domains that we have identified in both the domain and symptoms clusterings. On the other hand, the study has several notable strengths: (1) the sample size, which to our knowledge is the largest international sample in this kind of study; (2) inclusion of patients in all disease stages; and (3) the use of detailed assessments both for motor and non-motor symptoms.

In conclusion, we present statistical confirmation of the growing recognition of NMS-dominant presentation of PD and its heterogeneity as shown in the clusters. The specific expression of several NMS domains such as mood/anxiety, sleep/fatigue, cognition, and urinary function suggests differential patterns of neurodegeneration involving non-dopaminergic pathways, as suggested from neuropathology studies by several authors (Jellinger 2012). The clinical recognition of subtypes as reflected by the domains clusters allow for treatments to be tailored and could be the beginning of subtype-specific treatment packages for PD (Marras and Chaudhuri 2016). In the future, clinical characterisation of these groups will allow for studies of natural history of the various non-motor dominant clusters identified in this paper. Translating results to clinical management or experimental designs would require, using the same database, the identification of inclusion and exclusion criteria of patients into specific subgroups.

# Acknowledgments

We gratefully acknowledge data collection efforts by all contributors, collaborators, and administrative staff related to the NILS global non-motor symptoms of PD natural history study, the scale development group of the PD non-motor symptoms scale, and the MDS non motor study group. KRC also acknowledges the NIHR biomedical research centre at the Institute of Psychiatry, Psychology and Neurosciences, Kings College, London, and the National Parkinson Foundation, USA.

# Funding

JM is supported by a Barry M. Goldwater scholarship. CB and PL have been partially supported by the Spanish Ministry of Economy and Competitiveness through the Cajal Blue Brain (C080020-09; the Spanish partner of the Blue Brain initiative from EPFL) and TIN2016-79684-P projects, by the Regional Government of Madrid through the S2013/ICE-2845-CASI-CAM-CM project, and by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 720270.

# References

Burn DJ, Landau S, Hindle JV, Samuel M, Wilson KC, Hurt CS, Brown RG; PROMS-PD Study Group. Parkinson’s disease motor subtypes and mood. Mov Disord 2012;27(3):379–86.

Chaudhuri KR, Martinez-Martin P, Brown RG, Sethi K, Stocchi F, Odin P *et al*. The metric properties of a novel non-motor symptoms scale for Parkinson’s disease: results from an international pilot study. Mov Disord 2007;22(13):1901–11.

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(8):e70244.

Flensborg Damholdt M, Shevlin M, Borghammer P, Larsen L, Ostergaard K. Clinical heterogeneity in Parkinson’s disease revisited: a latent profile analysis. Acta Neurol Scand 2012;125(5):311–8.

Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson’s disease. J Neurol Neurosurg Psychiatry 1988;51(6):745–52.

Gjerløff T, Fedorova T, Knudsen K, Munk OL, Nahimi A, Jacobsen S, et al. Imaging

acetylcholinesterase density in peripheral organs in Parkinson's disease with 11C-donepezil PET. Brain 2015;138(Pt 3):653–63.

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(1):10–20.

Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. Neurology 1967;17(5):427–42.

Hubert L, Arabie P. Comparing partitions. J Classification 1985;2(1):193–218.

Jellinger KA. Neuropathology of sporadic PD disease: evaluation and change of concepts. Mov Disord 2012;27:8–30.

Lees AJ, Hardy J, Revesz T. Parkinson’s disease. Lancet 2009;373(9680):2055–66.

Ma LY, Chan P, Gu ZQ, Li FF, Feng T. Heterogeneity among patients with Parkinson's disease: Cluster analysis and genetic association. J Neurol Sci 2015;351(1):41–45.

Marinus J, Visser M, Stiggelbout AM, Rabey JM, Martinez-Martin P, Bonuccelli U, *et al*. A short scale for the assessment of motor impairments and disabilities in Parkinson’s disease: the SPES/SCOPA. J Neurol Neurosurg Psychiatry 2004;75(3):388–95.

Marras C, Chaudhuri KR. Non-motor features of Parkinson’s disease subtypes. Mov Disord 2016;31(8):1095–102.

Martinez-Martin P, Benito-Leon J, Burguera JA, Castro A, Linazasoro G, Martinez-Castrillo JC, *et al*. The SCOPA-Motor Scale for assessment of Parkinson’s disease is a consistent and valid measure. J Clin Epidemiol 2005;58(7):674–9.

Martinez-Martin P, Forjaz MJ, Cubo E, Frades B, de Pedro Cuesta J; ELEP Project Members. Global versus factor-related impression of severity in Parkinson’s disease: a new clinimetric index (CISI-PD). Mov Disord. 2006;21(2):208–14.

Martinez-Martin P, Rodriguez-Blazquez C, Abe K, Bhattacharyya KB, Bloem BR, Carod-Artal FJ, *et al*. International study on the psychometric attributes of the Non-Motor Symptoms Scale in Parkinson disease. Neurology 2009a;73(19):1584–91.

Martinez-Martin P, Rodriguez-Blazquez C, Forjaz MJ, de Pedro J; Spanish-American Longitudinal PD Patient Study Group. The Clinical Impression of Severity Index for Parkinson’s Disease: international validation study. Mov Disord. 2009b;24(2):211–7.

Mollenhauer B, Zimmermann J, Sixel-Döring F, Focke NK, Wicke T, Ebentheuer J, et al. Monitoring of 30 marker candidates in early Parkinson’s disease as progression markers. Neurology 2016;87(2):168–77.

Pont-Sunyer C, Hotter A, Gaig C, Seppi K, Compta Y, Katzenschlager R, *et al*. The onset of nonmotor symptoms in Parkinson's disease (the ONSET PD study). Mov Disord 2015;30(2): 229–37.

Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, *et al*. MDS clinical diagnostic criteria for Parkinson’s disease. Mov Disord 2015;30(12):1591–601.

Ray Chaudhuri K, Rojo JM, Schapira AH, Brooks DJ, Stocchi F, Odin P, *et al*. A proposal for a comprehensive grading of Parkinson’s disease severity combining motor and non-motor assessments: meeting an unmet need. PLoS ONE 2013;8(2):e57221.

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(5):379–82.

Romenets SR, Gagnon JF, Latreille V, Panniset M, Chouinard S, Montplaisir J, *et al*. Rapid eye movement sleep behavior disorder and subtypes of Parkinson’s disease. Mov Disord 2012;27(8):996–1003.

Rosenberg A, Hirschberg J. V-measure: A conditional entropy-based external cluster evaluation measure. Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning 2007:410–20.

Sauerbier A, Jenner P, Todorova A, Chaudhuri KR. Non motor subtypes and Parkinson’s disease. Parkinsonism Relat Disord 2016;22(Suppl 1):S41–6.

Seichepine DR, Neargarder S, Miller IN, Riedel TM, Gilmore GC, Cronin-Golomb A. Relation of Parkinson’s disease subtypes to visual activities of daily living. J Int Neuropsychol Soc 2011;17(5):841–52.

Selikhova M, Williams DR, Kempster PA, Holton JL, Revesz T, Lees AJ. A clinico-pathological study of subtypes in Parkinson’s disease. Brain 2009;132(Pt 11):2947–57.

Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. JR Stat Soc Series B Stat Methodol 2001;63(2):411–23.

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(1):51–8.

Warren JD, Rohrer JD, Schott JM, Fox NC, Hardy J, Rossor MN. Molecular nexopathies: a new paradigm of neurodegenerative disease. Trends Neurosci 2013;36:561-569.

# Supplementary material

*Supplementary Figure 1.* **Gap statistics.** Plot of the gap statistic versus number of clusters with k-means on 500 bootstrapped samples of a) the domains clustering, and b) the symptoms clustering. Error bars represent standard error (). Per the method described in Tibshirani *et al*. (2001), the optimal number of clusters is the smallest k such that For the domains clustering, ; for the symptoms clustering, . The gap statistic for the optimal and the comparison to are marked with dotted lines.



*Supplementary Figure 2.* **Domains clustering boxplots.** Boxplots for domains clustering for each symptom and cluster.



*Supplementary Figure 3.* **Correlation with PD duration.** Correlation of applicable variables with PD duration. PD onset has a negative correlation with PD duration but is excluded since it provides no information about PD progression.



*Supplementary Figure 4.* **Symptoms against PD duration.** For clarity, scatterplot points are colored according to cluster and jittered. Smoothed loess curves for each cluster are drawn in their respective cluster colors. The black curve is the curve for the entire population, and the global mean score is marked with a dotted line.