Parkinson’s disease subtypes

Authors, affiliations

Corresponding author postal and email address

Running Title: Parkinson's disease subtypes

Abstract (< 400 words)

Keywords: a, b, c, d, e

Introduction

Parkinson's disease has been classically considered a motor disorder with rest tremor , rigidity and bradykinesia as the central core feature. However, the concept of Parkinson's disease has changed considerably in the last few years now prompting a revision of the diagnostic criteria of PD so as to include non motor symptoms in the core parameters. REF (Marras and Chaudhuri 201; Postuma et al 2015). While motor symptoms remain as part of the core diagnostic criteria, NMS now are recognized to predate the motor syndrome for up to 20 years (Hawkes 2016)) and underpin the prodromal stage of PD (Berg et al 2015). The range of NMS in PD have also been classified and are recognized as dopaminergic, non dopaminergic, fluctuating as well drug induced and genetic (Todorova et al 2014). NMS mostly arise from the multipeptid dysfunction that is now recognized in PD, in addition to the fact that the pathophysiology involves the central nervous system as well as many extra cerebral organs including the gut and the heart (Jellinger 2012). Cluster analysis based studies have already identified several "non motor "clusters in the presentation of PD (Erro et al 2013: Onset PD 2015) while clinical observational studies suggest existence of robust and specific nonmotor subtypes (Sauerbier et al 2015; Marras and Chaudhuri 2016). These observations show the remarkable variability in the expression of PD. The progression, evolution and journey of these NMS have been poorly studied and appear to be variable from one patient to another, while the presence or accumulation of symptoms may differ broadly between individuals (Thenganatt and Jankovic, 2014).

This variability in the course and expression of the disease has promoted the search of specific subtypes () whose identification could provide knowledge on their pathophysiology, tailored treatment, progression rate, and prognosis.

Several subtypes of PD have been identified, being the motor subtypes dominantly classified by Jankovic et al, Scheiss et al, and Eggert et al. Nonetheless, studies have showed these phenotypes to be unstable over time.

However, non-motor symptoms are an important part of the disorder and several studies have explored PD subtypes considering motor subtypes and their association with several 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; Damholdt et al., 2012), REM sleep behavior disorder (Romenets et al., 2012); and visual daily activities (Seichepine et al. 2011).

Subtypes exclusively based on NMS have also been proposed (Sauerbier et al., 2016), but few studies have been carried out exploring the combination of the motor disturbances and a wide diversity of NMS for obtaining a more complete picture of the reality in regard to the potential existence of subtypes characterized by a specific mixture of motor and non-motor manifestations (van Rooden et al., 2011; Erro et al., 2013; Fereshtehnejad et al., 2015; Ba et al., 2016), with prognostic connotations (de Lau et al., 2014).

The present study was aimed at searching PD subtypes on a wide database of PD patients using the combination of motor cardinal features (bradykinesia, rigidity, tremor, axial signs) and NMS features assessed using specific rater-based scales and techniques of cluster analysis.

Materials and Methods

Experimental Design

Combination of data from two independent international studies: the validation study of the Non-Motor Symptoms Scale (NMSS) ($n = 411$) (Martinez-Martin et al., 2009a) and baseline data from the Non-motor International Longitudinal Study (NILS) ($n = 540$) (on going) sharing most of the protocol and assessments (Ray Chaudhuri et al., 2013).

Patients

Consecutive patients, both genders, any disease stage, diagnosed with PD by a neurologist with competence in movement disorders and according to internationally recognized criteria (Gibb and Lees, 1988; Lees et al., 2009). 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 assessment of PD manifestations; and inability for giving informed consent. Patient recruitment was carried out across 15 countries in America, Asia, and Europe from 2007 to 2011.

Assessments

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 (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 theoretical NMSS total score vary from 0 to 360 (Chaudhuri et al., 2007; Martinez-Martin et al., 2009a), a Fig. representing the non-motor "symptomatic burden" (NMSB).
3. The original Hoehn and Yahr (HY) (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 running from 0 to 24 (Martinez-Martin et al., 2006; Martinez-Martin et al., 2009b).

Ethical aspects

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

Data analysis

Descriptive statistics (central tendency and dispersion, percentages) were applied to inform on the main characteristics of the sample. SCOPA-Motor Examination items were grouped for obtaining an estimation of the PD cardinal features scores: tremor (sum of items 1 and 2); bradykinesia (item 3); rigidity (item 4); and axial (sum of items 5 to 10).

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. 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 **???** (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 "domains clustering" on the nine aggregate nonmotor symptom domains and the four motor symptoms, and the second "symptoms clustering" on the 30 individual nonmotor symptoms only. Average-linkage hierarchical agglomerative clustering on the 30 nonmotor symptoms was also performed to observe how the symptoms group together.

Determining *k*

Various formal measures were used to determine the optimal number of clusters for the dataset. The optimal k according to the Gap Statistic and the 1-standard-error method suggested by Tibshirani (2001) was $k = 4$ (Supplementary Fig. 1). Other cluster determination methods (within sum of squares error scree plot, minimum average silhouette width) suggested $k = 2, 3, 4$, where $k = 2, 3$ simply divided the data uninformatively into groups with varying levels of overall PD severity. Thus $k = 4$ was selected to offer a good blend of model fit, informativeness, and parsimony.

Interpretation

For the domains clustering, we displayed the distribution of each symptom for the four clusters using boxplots, which allowed us to visualize the center and spread of each cluster. Since the number of variables was substantially larger for the symptoms clustering, we visualized results for the second analysis with a heatmap. Finally, for the hierarchical clustering on the symptoms themselves, we displayed results in a dendrogram and clustered the symptoms into four interpretable clusters.

For each symptom in both clusterings, we used one-way ANOVA and χ^2 tests to respectively check the equality of symptom means and proportions across the clusters found, using Bonferroni correction for multiple testing with corrected $p < 0.05$ considered significant. Differences among pairwise clusters were tested post-hoc using Tukey's range test for continuous means, or pairwise χ^2 tests for proportions, again with Bonferroni correction.

To compare the clusterings, we visualized cluster alignment with stacked barplots, 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 symptom severity and disease duration, we computed the correlation coefficient for each symptom with PD duration and fitted smoothed loess curves to the data both globally and for each cluster found in the previous analyses.

Results

There were not significant differences between the excluded and included groups in regard to sex, PD duration, and HY ($\chi^2 \geq 0.19$), being difference in age in the limit of statistical significance (67.36 ±10.26 vs. 64.28 ±9.86; Mann-Withney, $p = 0.051$). The characteristics of the sample included for analysis ($n = 904$) are displayed in Table 1.

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
SCOPA-Motor Total score	21.07	12.06	19	1 – 72
A. Examination	11.54	6.56	10	0 – 41
B. Activities of daily living	6.80	4.19	7	0 – 21
C. Complications	2.73	3.01	2	0 – 12
Cardinal features scores*				
Tremor	2.59	2.58	2	0 – 12
Bradykinesia	2.40	1.41	2	0 – 6
Rigidity	2.23	1.36	2	0 – 6
Axial	3.25	2.67	3	0 – 12
Non-Motor Symptoms Scale total score	50.45	41.72	39	0 – 225

Cardiovascular	1.74	3.26	0	0 – 24
Sleep/Fatigue	8.76	8.71	6	0 – 48
Mood/Apathy	8.67	11.54	4	0 – 60
Perceptual problems/Hallucinations	1.64	3.86	0	0 – 33
Attention/Memory	5.40	7.42	2	0 – 36
Gastrointestinal	5.53	6.78	3	0 – 36
Urinary	8.07	8.93	5	0 – 36
Sexual function	3.53	5.98	0	0 – 24
Miscellaneous	7.12	7.78	4	0 – 48
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.

Patients were predominantly males (62.17%) and 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.

Domains clustering

Results from the clustering on the nine nonmotor domains and the four motor symptoms, as well as additional variables not used in the analysis, are reported in Table 2, with boxplots in Fig. 1. Cluster means for all variables were found to be statistically different, except for PD onset and sex ($p < 0.05/19 \approx 0.0026$ after correcting for the comparisons of the 19 variables). Specific pairwise differences are noted in the table.

Cluster 1 ($n = 428$) patients were mildly affected in all domains. This cluster was characterized by relatively lower disease durations and ages.

Cluster 2 ($n = 180$) patients were severely affected in nonmotor domains but mildly affected in motor domains. This cluster had an incidence of motor symptoms relatively similar to the cluster 1 (mild) subtype especially in tremor, but generally expressed significantly higher scores for nonmotor symptoms than clusters 1 and 3, especially in the sleep/fatigue, mood/apathy and miscellaneous domains. In motor complications uniquely, symptom scores were not statistically different from cluster 3.

Cluster 3 ($n = 232$) patients were severely affected in motor domains but mildly affected in nonmotor 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 scores of clusters 2 and 3 were not statistically significantly different.

Cluster 4 ($n = 64$) 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 oldest ages, but did not have a significantly different age of PD onset.

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%)
Nonmotor domains	Cardiovascular	0.7 (1.5)[24]	3.2 (3.7)[134]	1.1 (2.1)[24]	6.9 (6.4)[123]
	Sleep/fatigue	4.5 (5.0)[234]	16 (8.7)[134]	7.5 (6.6)[124]	21.7 (9.7)[123]
	Mood/apathy	3.4 (4.8)[234]	19.2 (15.0)[13]	6.6 (8.0)[124]	21.7 (13.5)[13]
	Perception/hallucination	0.5 (1.7)[24]	2.7 (4.3)[134]	0.8 (1.8)[24]	9.7 (6.9)[123]
	Attention/memory	3.0 (4.5)[24]	10.5 (9.2)[134]	3.3 (4.4)[24]	14.5 (11.0)[123]
	Gastrointestinal	2.9 (4.1)[234]	8.5 (7.1)[134]	4.7 (5.3)[124]	17.4 (9.2)[123]
	Urinary	4.7 (6.2)[24]	14.0 (9.9)[134]	6.2 (6.7)[24]	20.3 (9.7)[123]
	Sexual function	1.7 (3.4)[234]	7.3 (7.8)[13]	2.4 (4.1)[124]	9.0 (9.9)[13]
	Miscellaneous	4.0 (4.8)[234]	13.2 (8.7)[13]	6.2 (6.8)[124]	14.5 (10.1)[13]
Motor symptoms	Tremor	2.0 (1.9)[34]	1.5 (1.8)[34]	4.1 (2.8)[12]	4.1 (4.1)[12]
	Bradykinesia	1.6 (0.9)[234]	2.2 (1.1)[134]	3.5 (1.0)[124]	4.5 (1.3)[123]
	Rigidity	1.5 (0.9)[234]	1.8 (1.2)[134]	3.3 (1.0)[124]	4.2 (1.2)[123]
	Axial	1.7 (1.5)[234]	3.6 (2.2)[134]	4.5 (2.3)[124]	8.2 (2.7)[123]
	Motor complications	1.4 (2.1)[234]	3.1 (2.9)[14]	3.7 (2.9)[14]	7.0 (3.6)[123]
Variables not used in analysis	Sex (% male)	64	54	67	58
	CISI-PD total	5.5 (3.0)[234]	9.6 (3.8)[14]	10.1 (3.5)[14]	16.4 (4.6)[123]
	Age	62.5 (9.7)[4]	65.2 (9.4)[4]	64.9 (10.1)[4]	71.1 (7.9)[123]
	PD onset	56 (10.5)	56.6 (10.6)	56.3 (11.3)	56.7 (10.6)
	PD duration	6.5 (4.7)[234]	8.6 (5.7)[14]	8.6 (5.7)[14]	14.4 (8.0)[123]

[1] Significant difference with cluster 1 (corrected $p < 0.05$)

[2] Significant difference with cluster 2 (corrected $p < 0.05$)

[3] Significant difference with cluster 3 (corrected $p < 0.05$)

[4] Significant difference with cluster 4 (corrected $p < 0.05$)

Symptoms clustering

k-means clustering performed on the 30 individual nonmotor symptoms found slightly similar patterns to the first clustering (Table 2, heatmap in Fig. 2). Means of all symptoms were found to differ across clusters except for PD onset, sex, and tremor. Clusters 1 ($n = 406$) and 4 ($n = 88$), the mild and severe subtypes of the previous clustering, retained similar characteristics here; specifically, cluster 4 demonstrated especially high severity in the perception/hallucination and attention/memory domains. However, clusters 2 ($n = 189$) and 3 ($n = 221$) were characterized differently. Cluster 2 was strictly a mood/apathy-dominant cluster, with a relatively higher proportion of female patients. Cluster 3 did not exhibit any especially strong tendencies in any symptoms, and was thus classified as a group of patients with average severity.

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

	Cluster	1	2	3	4
	n (%)	406 (45%)	189 (21%)	221 (24%)	88 (10%)
1. Cardiovascular	Lightheadedness	0.6 (1.3)[234]	1.9 (2.6)[14]	2.1 (3)[14]	5 (4.2)[123]
	Fainting	0.1 (0.7)[4]	0.4 (1.1)[4]	0.1 (0.6)[4]	2 (3.3)[123]
2. Sleep/fatigue	Drowsiness	1 (2)[234]	2.3 (3.1)[14]	2.5 (2.9)[14]	6.4 (3.9)[123]
	Fatigue	1.3 (2)[234]	5.8 (4.3)[14]	4.9 (3.8)[14]	7.7 (3.7)[123]
	Insomnia	1.2 (2.4)[234]	4.9 (4.7)[13]	3.4 (4)[12]	4.9 (4.2)[1]
	RLS	0.7 (1.7)[234]	2 (3.2)[14]	2 (3.3)[14]	3.6 (3.9)[123]
3. Mood/apathy	Loss_interest	0.5 (1.2)[24]	6 (3.9)[134]	0.8 (1.4)[24]	4.7 (3.5)[123]
	Loss_activities	0.8 (1.6)[234]	7 (4)[134]	1.5 (2.4)[124]	5.5 (4.2)[123]
	Anxiety	1 (1.8)[234]	6.1 (4.5)[13]	1.8 (2.7)[124]	5.4 (4.1)[13]
	Depression	0.9 (1.7)[234]	7.5 (3.9)[134]	2.1 (2.8)[124]	5.3 (4)[123]
	Flat_affect	0.4 (1.1)[24]	5 (4.4)[124]	0.7 (1.6)[24]	2.9 (3.1)[123]
	Loss_pleasure	0.4 (1.2)[24]	6.9 (4)[134]	0.9 (1.8)[24]	3.7 (3.6)[123]
4. Perception/ hallucination	Hallucination	0.2 (1.1)[4]	0.8 (2.2)[4]	0.5 (1.4)[4]	3.7 (3.6)[123]
	Delusion	0.1 (0.7)[24]	1.1 (2.7)[134]	0.1 (0.9)[24]	3.1 (3.8)[123]
	Diplopia	0.2 (0.8)[24]	1 (2.5)[14]	0.5 (1.4)[4]	3.4 (4.1)[123]
5. Attention/ memory	Loss_concentration	0.9 (1.8)[234]	3.6 (3.4)[134]	2.4 (3.1)[124]	6.6 (3.8)[123]
	Forget_explicit	0.9 (1.6)[234]	2.6 (3.1)[14]	2.3 (2.9)[14]	7.2 (3.6)[123]
	Forget_implicit	0.7 (1.5)[234]	2.2 (3)[14]	1.9 (2.7)[14]	6.7 (4.1)[123]
6. Gastrointestinal	Drooling	0.7 (1.7)[234]	2.9 (4.1)[14]	3 (3.8)[14]	5.2 (4.6)[123]

	Swallowing	0.3 (1.1)[234]	1.2 (2.4)[14]	1.8 (2.7)[14]	4 (3.8)[123]
	Constipation	1.6 (2.8)[234]	3.6 (4.3)[14]	3.6 (4.2)[14]	6 (5.1)[123]
7. Urinary	Urinary_urgency	0.9 (1.7)[234]	3 (3.8)[14]	4.3 (3.8)[14]	7.2 (4.2)[123]
	Urinary_frequency	0.9 (1.7)[234]	3.3 (3.6)[14]	4.2 (4.1)[14]	6.6 (4.2)[123]
	Nocturia	1.6 (2.2)[234]	4.5 (4.1)[14]	5.5 (4.1)[14]	7.5 (4)[123]
8. Sexual	Sex_drive	0.7 (1.8)[234]	3.9 (4.6)[1]	2.6 (3.8)[14]	5.1 (5.3)[13]
	Sex_dysfunction	0.7 (2)[234]	2.4 (4.2)[14]	2.9 (4)[14]	4.7 (5.4)[123]
9. Miscellaneous	Unexplained_pain	0.9 (2.2)[234]	2.9 (4.1)[1]	2.4 (3.8)[14]	4.2 (4.2)[13]
	Gustation_olfaction	1.4 (2.8)[234]	3.5 (4.1)[14=]	3.3 (4)[1]	4.8 (4.8)[1]
	Weight_change	0.9 (1.7)[234]	2 (2.9)[14]	2 (3.1)[14]	4 (4.3)[123]
	Sweating	0.7 (1.9)[234]	2.1 (3.6)[1]	2.8 (4.1)[1]	3.1 (3.9)[1]
Motor symptoms	Tremor	2.6 (2.4)	2.3 (2.5)	2.4 (2.5)	3.4 (4)
	Bradykinesia	2.1 (1.3)[24]	2.8 (1.4)[1]	2.5 (1.4)[4]	3.5 (1.8)[13]
	Rigidity	2.0 (1.2)[4]	2.5 (1.3)[4]	2.3 (1.4)[4]	3.3 (1.7)[123]
	Axial	2.4 (2.1)[234]	4.4 (2.8)[14]	3.8 (2.6)[14]	6.5 (3.3)[123]
	Motor_comp	2.0 (2.5)[234]	3.5 (3.3)[14]	3.3 (3)[14]	5.4 (4)[123]
Variables not used in analysis	Sex (% male)	63	53[3]	71[2]	58
	CISI-PD total	6.6 (3.6)[234]	10.5 (4.5)[14]	9.6 (4.1)[14]	14.2 (5.9)[123]
	Age	62.9 (10.1)[34]	63.4 (9.7)[4]	66 (8.8)[1]	70.6 (8.3)[12]
	PD onset	56 (11)	55.1 (10.5)	56.8 (10.2)	58.3 (10.6)
	PD duration	6.9 (5)[34]	8.3 (5)[4]	9.2 (6.3)[14]	12.3 (8.2)[123]

[1] Significant difference with cluster 1 (corrected $p < 0.05$)

[2] Significant difference with cluster 2 (corrected $p < 0.05$)

[3] Significant difference with cluster 3 (corrected $p < 0.05$)

[4] Significant difference with cluster 4 (corrected $p < 0.05$)

Comparison between clusterings

The two clusterings were not very similar ($\text{ARI} = 0.37$; v-measure $= 0.37$). To help compare the alignment of the two clusterings, we denoted the four clusters in the domains clustering as D_1, D_2, D_3, D_4, and the clusters in the symptoms clustering as S_1, S_2, S_3, S_4. Alignment of the clusters is visualized in Fig. 3. While clusters D_1/S_1 and D_4/S_4 aligned well with each other, main differences in the clusterings occurred in the mixing of clusters 2 and 3; for example, D_3 (motor-dominant) was somewhat evenly split among S_1 (mild) and S_3 (average).

Hierarchical clustering on symptoms

Hierarchical clustering on the 30 nonmotor symptoms and the four motor symptoms produced the dendrogram in Fig. 4. Predictably, symptoms in the same nonmotor domains tended to cluster together, with some exceptions. Diplopia (domain 4) was grouped away from fellow perception/hallucination symptoms. RLS (domain 2) was grouped with miscellaneous domain 9 symptoms, albeit at a very high point on the tree, indicating a high level of dissimilarity. Similarly, drowsiness (domain 2) was grouped with the attention/memory symptoms of domain 5. Notably, tremor was the most dissimilar symptom, occupying a single branch at the top of the tree.

Longitudinal analysis

Most variables had little to no correlation with PD duration (Supplementary Fig. 2). Variables with the highest correlation included symptoms urinary frequency, swallowing, and drooling; domains miscellaneous, urinary, and gastrointestinal; and total CISI score. Scatterplots for CISI Total, Tremor, Anxiety, and Depression appear in Supplementary Fig. 3.

Discussion

Domains clustering reveals clusters that confirm previous findings in the field, mainly van Rooden et al. and the identification of four subtypes of Parkinson's disease: mild, nonmotor-dominant, motor-dominant, and severe. van Rooden's work was conducted with a separate dataset using a different modeling method (expectation-maximization), and this investigation independently confirms these subtype classifications. Unlike van Rooden, mean disease durations differences do exist between subtypes 1 (mild) and 4 (severe), likely due to further development of the disease, although the differences between 2 and 3 (nonmotor/motor-dominant) subtypes are insignificant (Table 2), importantly suggesting different developmental paths of the disease.

Overall, little information was found in PD onset, PD duration, or current age. Mean ages were similar for clusters 1, 2, and 3 ($p > 0.05$), but different from the severe cluster 4, intuitively since cluster 4 represented a more advanced stage of the disease. Specifically, clusters 1 and 4 seemed to be phenotypically quite similar, except at different stages of disease progression, given cluster 4's higher ages and disease durations.

However, clusters 2 and 3 clearly showed different disease progression, one in the motor direction, and one in the nonmotor. Both groups have similar age, PD onset, and PD duration

scores, but differ significantly in symptomatic expression. Cluster 2 is dominated by a high severity of nonmotor domains, especially Sleep/fatigue, Mood/apathy, Urinary, and Miscellaneous. Cluster 3, however, is dominated by a high prevalence of motor symptoms, where most motor symptoms are similar to the mild cluster 1. Of note is that the tremor population mean is the highest cluster mean, even higher than the severe subtype 4. This motor-dominant cluster may thus overlap with Ma et al.'s tremor dominant/slow progression cluster.

Generally, given stable PD onset scores and predictably increasing PD duration scores for clusters 1 and 4, Ma et al.'s rapid disease progression/late onset and tremor dominant/slow progression clusters were mostly not found in this dataset, save for the tremor-dominant motor cluster.

Our longitudinal analysis gives more insight into the clusters found in the previous analyses. According to Supplementary Fig. 2, most symptoms are uncorrelated with PD duration, especially Mood/apathy symptoms (anxiety, depression) and tremor. The differences in disease progression for each cluster can be seen by the corresponding graphs in Supplementary Fig. 3. Cluster 2 (Nonmotor-dominant) starts at higher scores for anxiety and depression, and actually decreases with increasing PD duration, thus indicating that these patients' subtype can be determined early after PD onset from the depressive symptom score. Similarly, when examining cluster 3 (Motor-dominant), the mean tremor score is substantially higher from PD onset. Interestingly, cluster 4 (Severe) generally starts at lower tremor and motor scores during disease, but then rises sharply, exceeding other clusters. More evidence that tremor is a unique motor symptom is located in Fig. 3, where it is the most distant symptom from all other symptoms.

It's intuitive that a Depression-Dominant group emerges from the symptoms clustering, since the Mood/apathy domain consists of 5 separate measures. Thus, any high expression of depressive symptoms is magnified in clustering, since the symptoms are highly similar (Fig. 3) and treated with equal weight. Once again reinforcing what was discovered previously, depressive symptoms have been shown to be very important in determining subtypes of PD.

Acknowledgments

Funding

Supplementary material

References

Ba F, Obaid M, Wieler M, Camicioli R, Martin WR. Parkinson's Disease: The relationship between non-motor symptoms and motor phenotype. Can J Neurol Sci. 2016;43(2):261-7.

Berg D, Postuma RB, Adler CH, Bloem BR, Chan P, Dubois B, et al. MDS research criteria for prodromal Parkinson's disease. Mov Disord. 2015;30(12):1600-11.

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.

de Lau LM, Verbaan D, van Rooden SM, Marinus J, van Hilten JJ. Relation of clinical subtypes in Parkinson's disease with survival. Mov Disord. 2014;29(1):150-1.

Eggert

Erro R

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.

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(8):863-73.

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.

Graham JM, Sagar HJ. A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes. MovDisord. 1999;14(1):10-20.

Hawkes CH, Del Tredici K, Braak H. A timeline for Parkinson's disease. Parkinsonism Relat Disord. 2010;16(2):79-84.

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

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

Jankovic J, McDermott M, Carter J, Gauthier S, Goetz C, Golbe L, et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. Neurology. 1990;40(10):1529-34.

Jellinger K

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

Liu P, Feng T, Wang YJ, Zhang X, Chen B. Clinical heterogeneity in patients with early-stage Parkinson's disease: a cluster analysis. J Zhejiang Univ Sci B. 2011;12(9):694-703.

Ma LY, Chan P, Gu ZQ, Li FF, Feng T. Heterogeneity among patients with Parksinon's disease: Cluster analysis and genetic association. Journal of the Neurological Sciences 2015;351(1):41–5.

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. Nonmotor features of Parkinson's disease subtypes. Mov Disord. 2016. doi: 10.1002/mds.26510.

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, Rodríguez-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.

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 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2007;410–20.

Sauerbier A

Sauerbier A, Jenner P, Todorova A, Chaudhuri KR. Non motor subtypes and Parkinson's disease. ParkinsonismRelatDisord. 2016;22Suppl 1:S41-6.

Scheiss

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.

Thenganatt MA, Jankovic J. Parkinson disease subtypes. JAMA Neurol. 2014;71(4):499-504.

Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2001;63(2):411–23.

Todorova A, Jenner P, Ray Chaudhuri K. Non-motor Parkinson's: integral to motor Parkinson's, yet often neglected. Pract Neurol. 2014;14(5):310-22.

van Rooden SM, Colas F, Martínez-Martín P, Visser M, Verbaan D, Marinus J, et al. Clinical subtypes of Parkinson's disease. Mov Disord. 2011;26(1):51-8.

Figure Legends

Figure 1. **Domains clustering boxplots.** Boxplots for domains clustering for each symptom and cluster.

Figure 2. **Symptoms clustering heatmap.** Heatmap of symptoms for each cluster, separated by white lines according to motor symptoms, nonmotor domains, and variables not included in the analysis. Since symptoms have different scales, cluster means for each symptom are displayed as standardized scores relative to the global symptom mean.

Figure 3. **Cluster alignment.** For each cluster in the domains and symptoms clustering, stacked bars indicate the cluster's composition according to the other clustering.

Figure 4. **Hierarchical clustering on symptoms.** Average-linkage hierarchical clustering of motor (blue) and nonmotor (black) symptoms. Symptoms are labeled with their name and corresponding domain number. Dendrogram colored with 5 clusters.

Supplementary Figure 1. **Gap statistic.** Plot of the gap statistic Gap(k) versus number of clusters with k-means on 500 bootstrapped samples of the domains clustering. Error bars represent ±1

standard error (se). Per the method described in Tibshirani,1 the optimal number of clusters is the smallest k such that $Gap(k) \geq Gap(k+1) - se_{k+1}$. In this case, $k = 4$.

Supplementary Figure 2. **Longitudinal correlation.** Correlation of each variable except sex and PD onset with PD duration.

Supplementary Figure 3. Selected symptoms plotted against PD duration. Smoothed loess curves for each cluster are drawn in their respective clusters. The black curve is the curve for the entire population, and the global mean score is marked with a dotted line.