Parkinson's disease cluster analysis

Abstract

Background This is the background.
Methods These are methods.
Findings These are findings.
Interpretation This is interpretation.
Funding This is funding.

1. Introduction

2. Methods

2.1. Statistical analysis

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 symptoms were scaled to z-scores ($\mu = 0$, $\sigma = 1$) before clustering, and unscaled afterwards for interpretation. Analyses were conducted in R 3.2.4 (www.r-project.org).

2.1.1. Cluster analysis

k-means was used for cluster analysis. We performed two analyses on the patients in the dataset: the first on the nine aggregate nonmotor symptom domains and the four motor symptoms ("domains clustering"), and the second on the 30 individual nonmotor symptoms collected with the four motor symptoms ("symptoms clustering"). Complete-linkage hierarchical clustering on the 30 nonmotor symptoms was also performed to observe the categorization of symptoms.

2.1.2. Determining k

Various formal measures were used to determine the optimal number of clusters for the dataset. First, the optimal k according to the Gap Statistic and the 1-standard-error method[1] was $k = 4$ (Figure 1). Second, although the globally-optimal clusters reported from stability measures of the clValid R package[2] stability were 2, 6, and 8 (Figure 2), there appeared to be consensus that $k = 4$ is a locally-optimal number of clusters which supported the gap statistic result.

Other cluster validation methods (SSE scree plot, minimum average silhouette width) generally suggested 2 clusters. As a consequence, $k = 2, 3, 4$ was tried, 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.

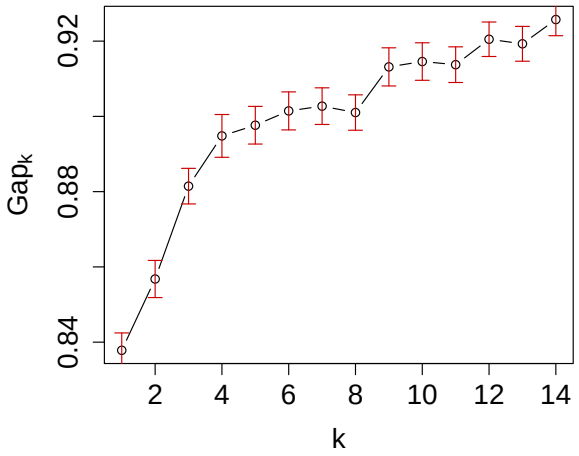

Figure 1: Plot of gap statistic versus number of clusters with k-means on 100 bootstrapped samples. Error bars are standard error. Per the method described in Tibshirani[1], 4 is the smallest k such that $\text{Gap}(k) \geq \text{Gap}(k+1) - \text{se}_{k+1}$.

2.1.3. 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 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 test 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 with Bonferroni

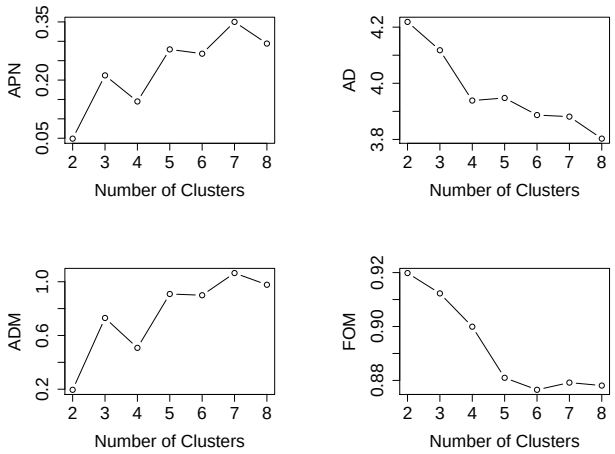

Figure 2: Average proportion of non-overlap (APN), average distance (AD), average distance between means (ADM), and figure of merit (FOM) stability measures[3,4]. Statistics should be minimized; notice the local optima at $k = 4$ for APN, AD, and ADM.

correction for proportions.

To compare the clusterings, we computed the adjusted rand index[5] (ARI) and v-measure[6] to evaluate similarity between the two clusterings, and used stacked barplots to visualize the alignment of the clusters.

Lastly, to explore the relationship between symptom severity and disease duration, we computed the linear correlation 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.

3. Results

3.1. Domains clustering

k-means clustering on the nine nonmotor domains and the four motor symptoms found four clusters (Table 1; boxplots in Figure 3). Cluster means for all symptoms were found to be statistically different ($p < 0.05/13 \approx 0.003$), with specific pairwise differences noted in the table. Additionally, statistics for variables not used in the clustering were explored similarly in Table 2; cluster means for all variables were found to be statistically different, except for PD onset.

Cluster 1 ($n = 406$) patients were mildly affected in all domains. This cluster was characterized by relatively lower disease durations and ages.

Cluster 2 ($n = 189$) 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/cognition and miscellaneous domains. This group also had a statistically higher percentage of females than Cluster 3.

Cluster 3 ($n = 221$) 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, but less than 4, with the exception of tremor, which was especially high.

Cluster 4 ($n = 88$) 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.

3.2. Symptoms clustering

k-means clustering performed on the 30 individual nonmotor symptoms found similar patterns to the first clustering (Table 3, heatmap in Figure 6). Means of all symptoms were found to differ across clusters except for PD_Onset and tremor.

Clusters 1 ($n = 509$) and 4 ($n = 49$), the mild and severe subtypes of the previous clustering, retained their characteristics here, while clusters 2 ($n = 97$) and 3 ($n = 249$) were different. 2 was a mood/cognition-dominant cluster, with a relatively higher proportion of female patients and additionally a more severe incidence of insomnia. 3 was not characterized by any especially high symptoms and was thus classified as a group of patients with average severity.

3.3. Comparison between clusterings

The two clusterings are somewhat similar (ARI $= 0.46$; v-measure $= 0.43$). Alignment of the clusters is visualized in Figure 4. While clusters 1 and 4 of both clusterings align well with each other, main differences in the clusters occur in the mixing of clusters 2 and 3 of both clusterings; for example, domains cluster 3 (motor-dominant) is somewhat evenly split among symptoms clusters 1 (mild) and 3 (average).

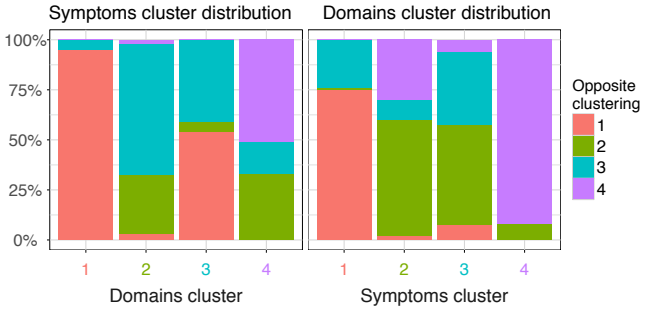

Figure 4: Bar chart indicating the distribution of opposite cluster membership for the patients of a given cluster. For example, the bar furthest left indicates that close to 100% of patients in domains cluster 1 are in symptoms cluster 1, with the rest in symptoms cluster 3.

Table 1: Summaries of domains clustering. Unless otherwise specified, statistics are reported as mean (sd). **Bold** indicates the highest score for that symptom out of the four clusters.

Cluster	1	2	3	4
n	406 (45%)	189 (21%)	221 (24%)	88 (10%)
Nonmotor (Domains)				
Cardiovascular (0–24)	0.7 (1.6)[24]	2.3 (2.9)[134]	1.1 (2.0)[24]	**7.0** (6.0)[234]
Sleep/fatigue (0–48)	4.3 (4.8)[234]	14.9 (8.2)[134]	6.9 (6.3)[124]	**21.0** (10.0)[123]
Mood/cognition (0–60)	3.1 (4.5)[234]	16.1 (14.0)[134]	6.6 (8.1)[124]	**23.4** (14.3)[123]
Perception/hallucination (0–36)	0.5 (1.7)[24]	1.9 (3.5)[134]	0.8 (2.0)[24]	**8.6** (6.9)[123]
Attention/memory (0–36)	2.8 (4.2)[24]	8.9 (8.1)[134]	3.2 (4.4)[24]	**15.6** (11.1)[123]
Gastrointestinal (0–36)	2.8 (3.9)[234]	8.8 (7.1)[134]	4.1 (4.7)[124]	**14.6** (9.6)[123]
Urinary (0–36)	4.5 (5.9)[234]	12.4 (9.7)[134]	6.1 (6.4)[124]	**20.3** (9.9)[123]
Sexual function (0–24)	1.5 (3.1)[24]	6.4 (7.4)[134]	2.5 (4.1)[24]	**9.0** (9.7)[123]
Miscellaneous (0–48)	3.9 (4.7)[24]	13.2 (8.6)[13]	5.2 (5.8)[24]	**14.0** (9.5)[13]
Motor				
Tremor	1.9 (1.8)[34]	1.6 (1.9)[34]	**4.3** (2.8)[124]	3.5 (3.8)[123]
Bradykinesia	1.5 (0.9)[234]	2.1 (1.1)[134]	3.6 (0.9)[124]	**4.0** (1.5)[123]
Rigidity	1.5 (0.9)[234]	1.8 (1.1)[134]	3.3 (0.9)[124]	**3.8** (1.4)[123]
Axial	1.8 (1.6)[234]	3.3 (2.1)[134]	4.3 (2.4)[124]	**7.4** (2.9)[123]

[1] Significant difference with cluster 1 (adjusted $p < 0.05$)
[2] Significant difference with cluster 2 (adjusted $p < 0.05$)
[3] Significant difference with cluster 3 (adjusted $p < 0.05$)
[4] Significant difference with cluster 4 (adjusted $p < 0.05$)

Table 2: Cluster statistics for variables not used in the analysis

Cluster	1	2	3	4
Sex (% male)	63	53[3]	71[2]	58
Age	62.7 (9.7)[34]	64 (9.0)[4]	65 (10.2)[14]	70.5 (8.9)[123]
PD_Onset	56 (10.6)	55.2 (10.1)	56.9 (11.1)	58.2 (11.4)
PD_Duration	6.7 (4.8)[234]	8.7 (5.7)[14]	8 (5.5)[14]	12.3 (8.1)[123]
CISI_Total	5.7 (3.2)[234]	9.2 (3.9)[14]	9.7 (3.5)[14]	14.8 (5.1)[123]

[1] Significant difference with cluster 1 (adjusted $p < 0.05$)
[2] Significant difference with cluster 2 (adjusted $p < 0.05$)
[3] Significant difference with cluster 3 (adjusted $p < 0.05$)
[4] Significant difference with cluster 4 (adjusted $p < 0.05$)

3.4. Hierarchical clustering on symptoms

Hierarchical clustering on the 30 nonmotor symptoms and the four motor symptoms produced the dendrogram in Figure 5. Predictably, symptoms in the same nonmotor domains tended to cluster together, with the exceptions of diplopia (domain 4), drowsiness (domain 2), and the symptoms in domain 9 (miscellaneous). Notably, tremor was the most dissimilar symptom, occupying a single branch at the top of the tree.

3.5. Longitudinal analysis

Most variables had little to no correlation with PD duration (Figure 7). Variables with the highest correlation include 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 Figure 8.

4. Discussion

Domains clustering reveals clusters that confirm previous findings in the field, mainly van Rooden et al.[7] 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 predominated) subtypes are insignificant (Table 1), importantly suggesting different developmental paths of the disease.

Overall, little information was found in pdonset, du-

rat_pd, or current age. Mean ages were similar for clusters 1, 2, and 3 ($p > 0.05$), but different for the severe cluster 4, intuitively since cluster 4 represents a more advanced stage of the disease. Specifically, clusters 1 and 4 seem to be phenotypically quite similar, except at different stages of disease progression, given cluster 4's higher age and durat_pd scores.

However, clusters 2 and 3 clearly show 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/Cognition, 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's tremor dominant/slow progression cluster[8].

Generally, given stable pdonset scores and predictably increasing durat_pd scores for clusters 1 and 4, Ma et al's rapid disease progression/late onset and tremor dominant/slow progression clusters[8] 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 Figure 7, most symptoms are uncorrelated with PD duration, especially mood/cognition symptoms (anxiety, depression) and tremor. The differences in disease progression for each cluster can be seen by the corresponding graphs in Figure 8. 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 Figure 5, where it is the most distant symptom from all other symptoms.

It's intuitive that a Depression-Dominant group emerges when clustering on nms{1-30}, since domain 3 consists of 5 separate measures. Thus, any high expression of depressive symptoms is magnified in clustering, since the symptoms are highly similar (Figure 5) 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.

5. Supplementary Material

Probably the large table goes here?

6. References

[1] 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.

[2] Brock G, Pihur V, Datta S, Datta S. clvalid: An r package for cluster validation. Journal of Statistical Software 2008;25(1):1–22.

[3] Datta S, Datta S. Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes. BMC bioinformatics 2006;7(1):397.

[4] Yeung KY, Haynor DR, Ruzzo WL. Validating clustering for gene expression data. Bioinformatics 2001;17(4):309–18.

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

[6] Rosenberg A, Hirschberg J. V-measure: A conditional entropy-based external cluster evaluation measure. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). 2007, p. 410–20.

[7] 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. Movement Disorders 2010;25(8):969–78.

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

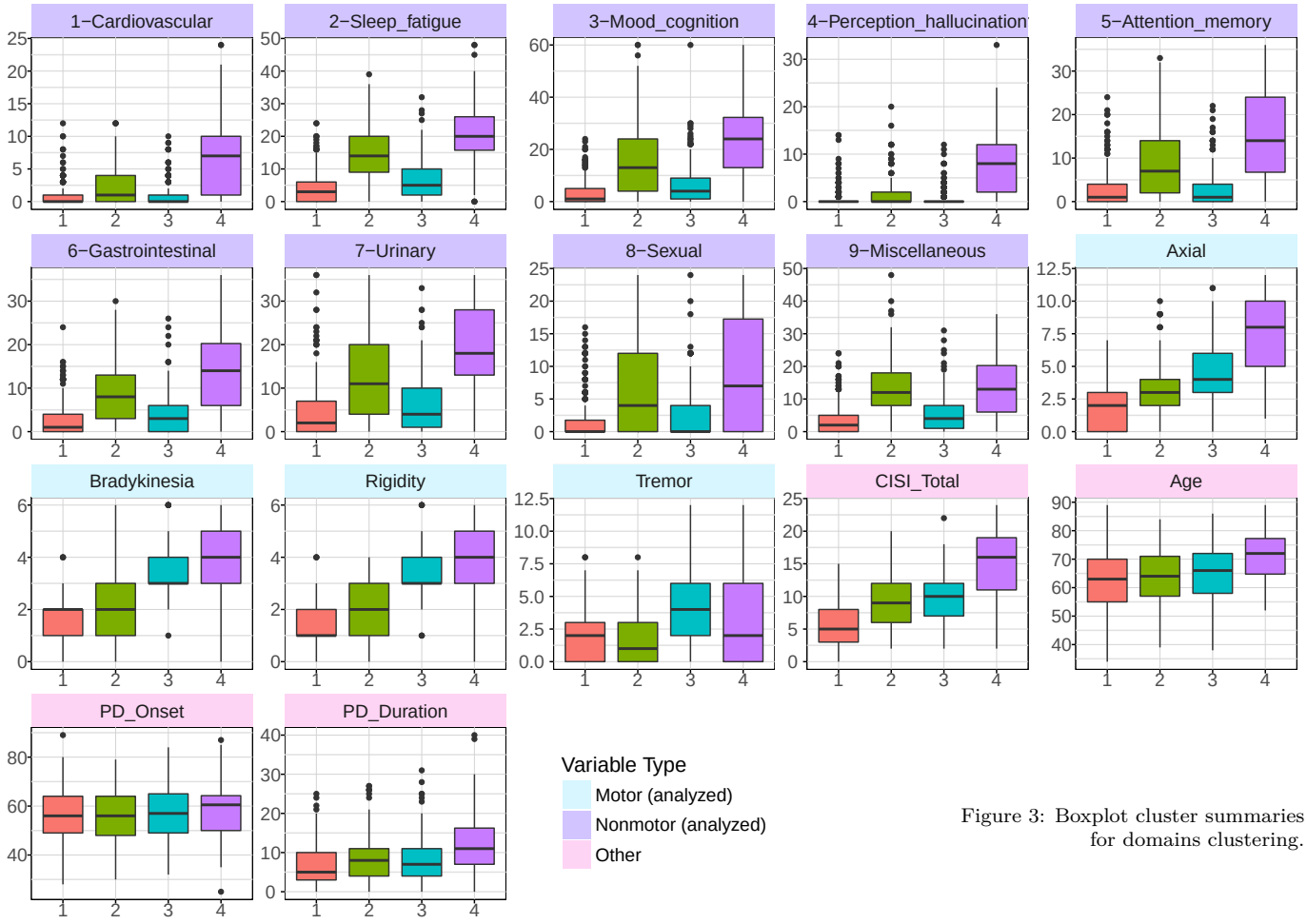

Figure 3: Boxplot cluster summaries for domains clustering.

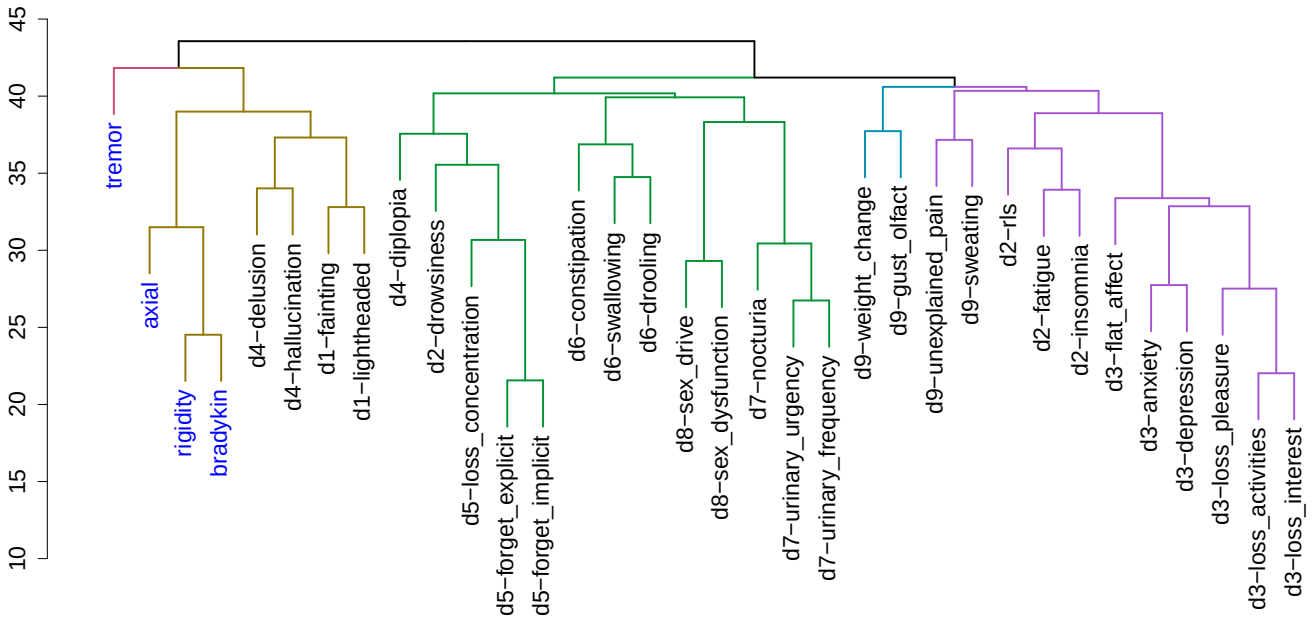

Figure 5: Hierarchical clustering of motor (blue) and nonmotor (black) symptoms. Symptoms are labeled with their name and corresponding domain number. Dendrogram colored with 4 clusters.

Table 3: Summaries of clustering on individual nonmotor domains. Unless otherwise specified, statistics are reported as mean (sd). **Bold** indicates the highest score for that symptom out of the four clusters.

	Cluster	1	2	3	4
	n	509 (56.3%)	97 (10.7%)	249 (27.5%)	49 (5.4%)
1. Cardiovascular	Lightheadedness	0.6 (1.3)^{234}	2.4 (3.0)^{14}	2.1 (2.9)^{14}	**5.2** (4.2)^{123}
	Fainting	0.1 (0.7)^{24}	0.5 (1.4)^{134}	0.2 (0.6)^{24}	**2.4** (3.5)^{123}
2. Sleep/fatigue	Drowsiness	1.0 (1.9)^{234}	2.8 (3.3)^{14}	2.5 (3.0)^{14}	**7.0** (3.8)^{123}
	Fatigue	1.3 (2.0)^{234}	6.3 (4.2)^{13}	4.8 (3.9)^{124}	**7.3** (3.5)^{13}
	Insomnia	1.2 (2.6)^{234}	**5.5** (4.8)^{13}	3.2 (3.8)^{12}	4.2 (3.9)^{1}
	RLS	0.7 (1.6)^{234}	2.3 (3.5)^{14}	1.9 (3.2)^{14}	**3.9** (3.9)^{123}
3. Mood/cognition	Loss_interest	0.4 (1.1)^{234}	**6.1** (3.8)^{134}	1.0 (1.8)^{124}	4.6 (3.7)^{123}
	Loss_activities	0.7 (1.6)^{234}	**7.1** (4.1)^{134}	1.7 (2.4)^{124}	5.4 (4.0)^{123}
	Anxiety	0.9 (1.8)^{234}	**6.6** (4.3)^{134}	2.0 (2.8)^{124}	4.6 (3.9)^{123}
	Depression	0.8 (1.6)^{234}	**7.8** (3.7)^{134}	2.3 (3.0)^{124}	4.3 (3.7)^{123}
	Flat_affect	0.4 (1.1)^{234}	**4.9** (4.3)^{134}	0.9 (1.9)^{124}	2.7 (3.0)^{123}
	Loss_pleasure	0.4 (1.2)^{234}	**6.8** (4.1)^{134}	1.1 (2.1)^{124}	3.4 (3.3)^{123}
4. Perception/ hallucination	Hallucination	0.2 (1.0)^{234}	0.8 (1.9)^{14}	0.6 (1.6)^{14}	**4.6** (3.6)^{123}
	Delusion	0.1 (0.7)^{24}	1.5 (3.1)^{124}	0.2 (1.0)^{24}	**3.1** (3.8)^{123}
	Diplopia	0.2 (0.7)^{234}	1.0 (2.5)^{14}	0.6 (1.4)^{14}	**4.4** (4.3)^{123}
5. Attention/ memory	Loss_concentration	0.9 (1.8)^{234}	4.0 (3.6)^{134}	2.5 (3.1)^{124}	**6.6** (3.9)^{123}
	Forget_explicit	0.9 (1.6)^{234}	3.8 (3.9)^{134}	2.2 (2.9)^{124}	**6.7** (3.6)^{123}
	Forget_implicit	0.7 (1.6)^{234}	3.1 (3.7)^{134}	1.8 (2.6)^{124}	**6.7** (4.0)^{123}
6. Gastrointestinal	Drooling	0.7 (1.6)^{234}	3.0 (4.1)^{14}	3.0 (3.8)^{14}	**6.1** (4.4)^{123}
	Swallowing	0.4 (1.2)^{234}	1.2 (2.1)^{14}	1.7 (2.8)^{14}	**4.5** (3.9)^{123}
	Constipation	1.6 (2.9)^{234}	3.6 (4.3)^{14}	3.3 (4.1)^{14}	**7.0** (5.1)^{123}
7. Urinary	Urinary_urgency	1.0 (1.8)^{234}	3.5 (4.1)^{14}	3.8 (3.8)^{14}	**7.8** (4.0)^{123}
	Urinary_frequency	0.9 (1.9)^{234}	4.0 (4.1)^{14}	3.8 (4.0)^{14}	**6.5** (4.0)^{123}
	Nocturia	1.7 (2.3)^{234}	5.0 (4.3)^{14}	5.1 (4.1)^{14}	**7.7** (3.9)^{123}
8. Sexual	Sex_drive	0.8 (1.9)^{234}	4.5 (4.9)^{13}	2.3 (3.6)^{124}	**5.2** (5.4)^{13}
	Sex_dysfunction	0.8 (2.2)^{234}	3.2 (4.7)^{1}	2.4 (3.8)^{14}	**4.6** (5.4)^{13}
9. Miscellaneous	Unexplained_pain	0.7 (1.9)^{234}	2.5 (3.7)^{14}	3.0 (4.1)^{14}	**4.2** (4.3)^{123}
	Gustation_olfaction	1.5 (2.9)^{234}	3.6 (4.3)^{1}	3.2 (3.9)^{14}	**4.7** (4.8)^{13}
	Weight_change	0.9 (1.7)^{234}	2.5 (3.4)^{14}	1.9 (3.0)^{14}	**3.9** (4.2)^{123}
	Sweating	0.7 (1.8)^{234}	2.2 (3.5)^{1}	2.8 (4.1)^{1}	**3.2** (3.8)^{1}
Motor	Tremor	2.5 (2.3)	2.1 (2.4)	2.6 (2.6)	**4.1** (4.3)
	Bradykinesia	2.0 (1.2)^{234}	2.8 (1.4)^{14}	2.8 (1.4)^{14}	**3.9** (1.8)^{123}
	Rigidity	1.9 (1.2)^{234}	2.4 (1.3)^{14}	2.6 (1.4)^{14}	**3.6** (1.6)^{123}
	Axial	2.2 (1.8)^{234}	4.4 (2.8)^{14}	4.2 (2.7)^{14}	**7.2** (3.2)^{123}
Variables not used in analysis	Age	62.7 (9.9)^{34}	64.5 (9.6)^{4}	66 (9.2)^{14}	71.5 (8.1)^{123}
	Sex (% Male)	65^{2}	51^{1}	60	63
	PD_Onset	55.9 (10.8)	56.2 (10.4)	56.7 (10.5)	57.8 (11.0)
	PD_Duration	6.8 (4.8)^{34}	8.3 (5.3)^{4}	9.3 (6.3)^{14}	13.8 (8.2)^{123}
	CISI_Total	6.3 (3.4)^{234}	10.8 (4.7)^{14}	10 (4.0)^{14}	15.1 (5.7)^{123}

[1] Significant difference with cluster 1 (adjusted $p < 0.05$)
[2] Significant difference with cluster 2 (adjusted $p < 0.05$)
[3] Significant difference with cluster 3 (adjusted $p < 0.05$)
[4] Significant difference with cluster 4 (adjusted $p < 0.05$)

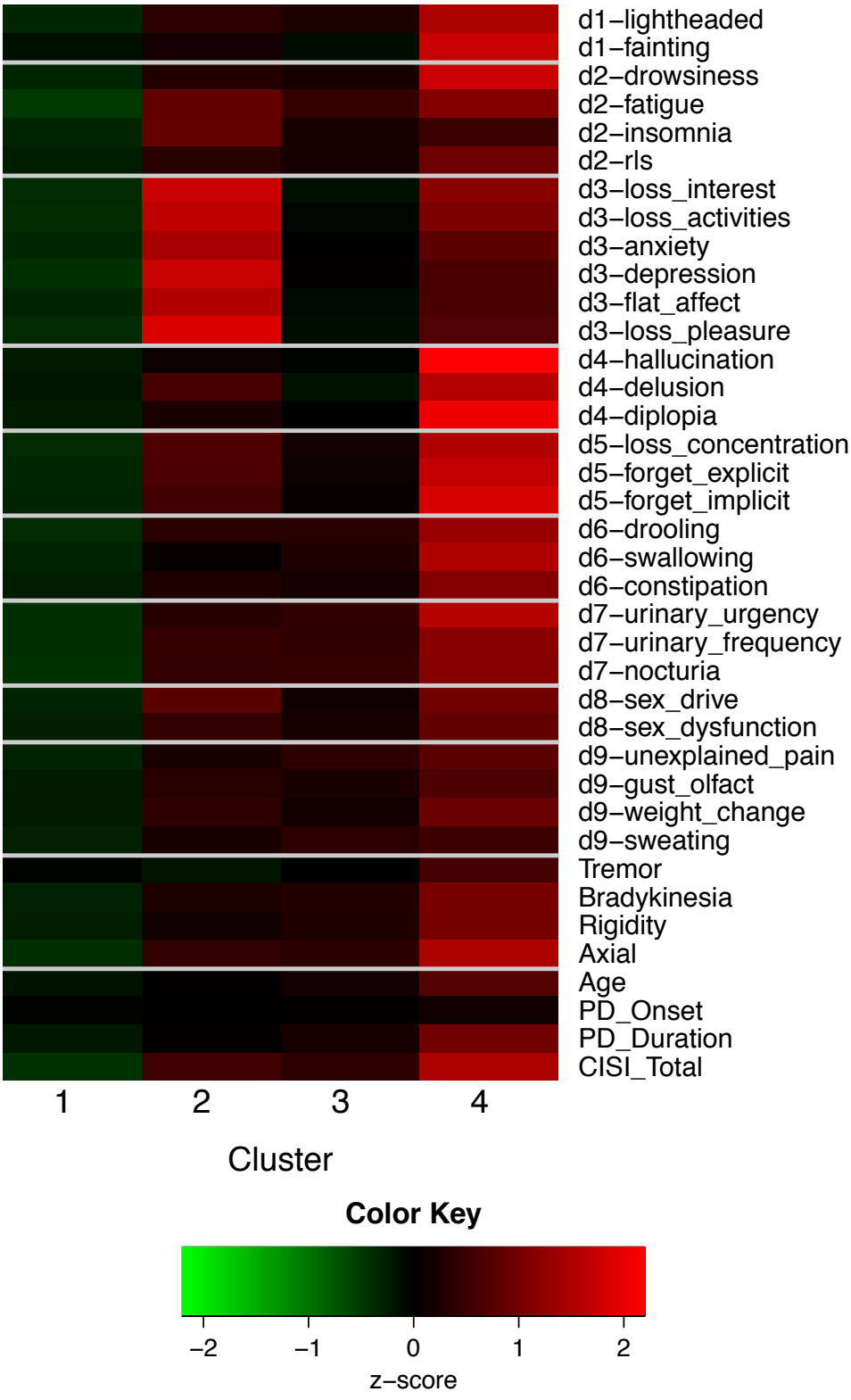

Figure 6: Heatmap of symptoms clustering. Color indicates cluster mean as a z-score for the given symptom.

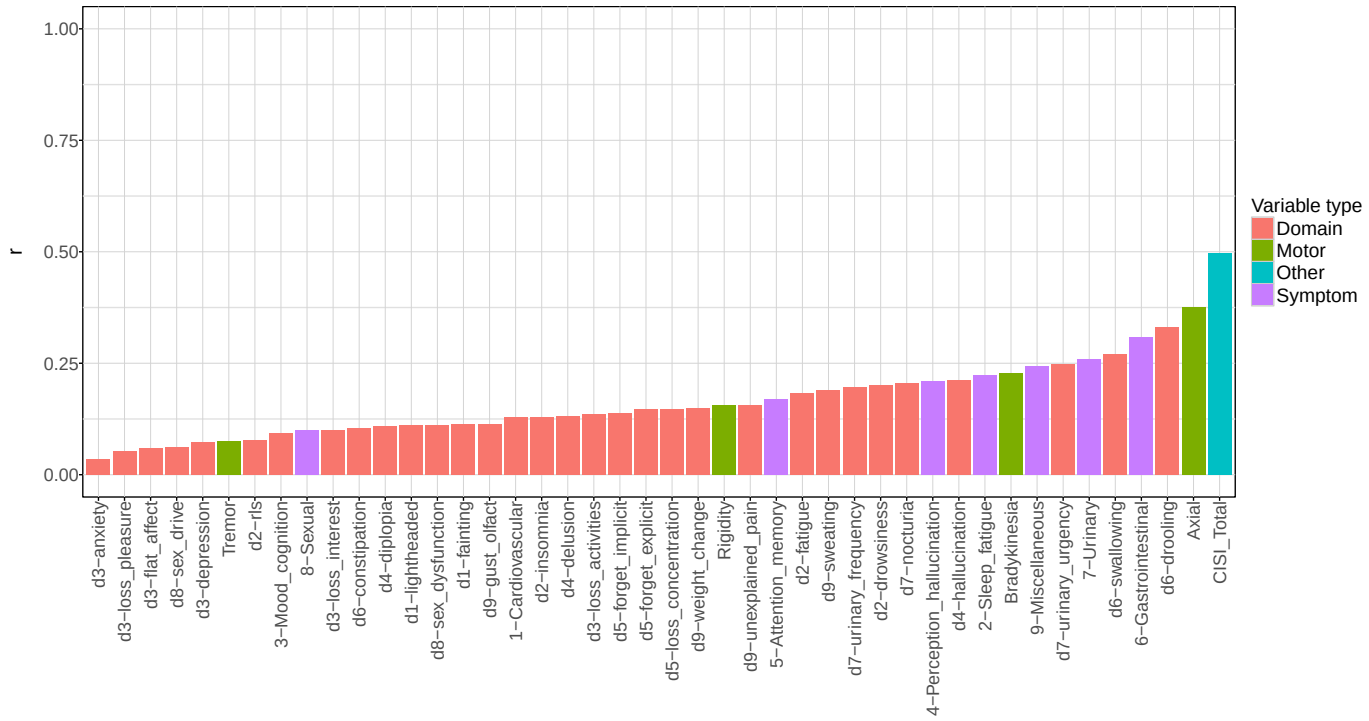

Figure 7: Linear correlation of each symptom with PD duration.

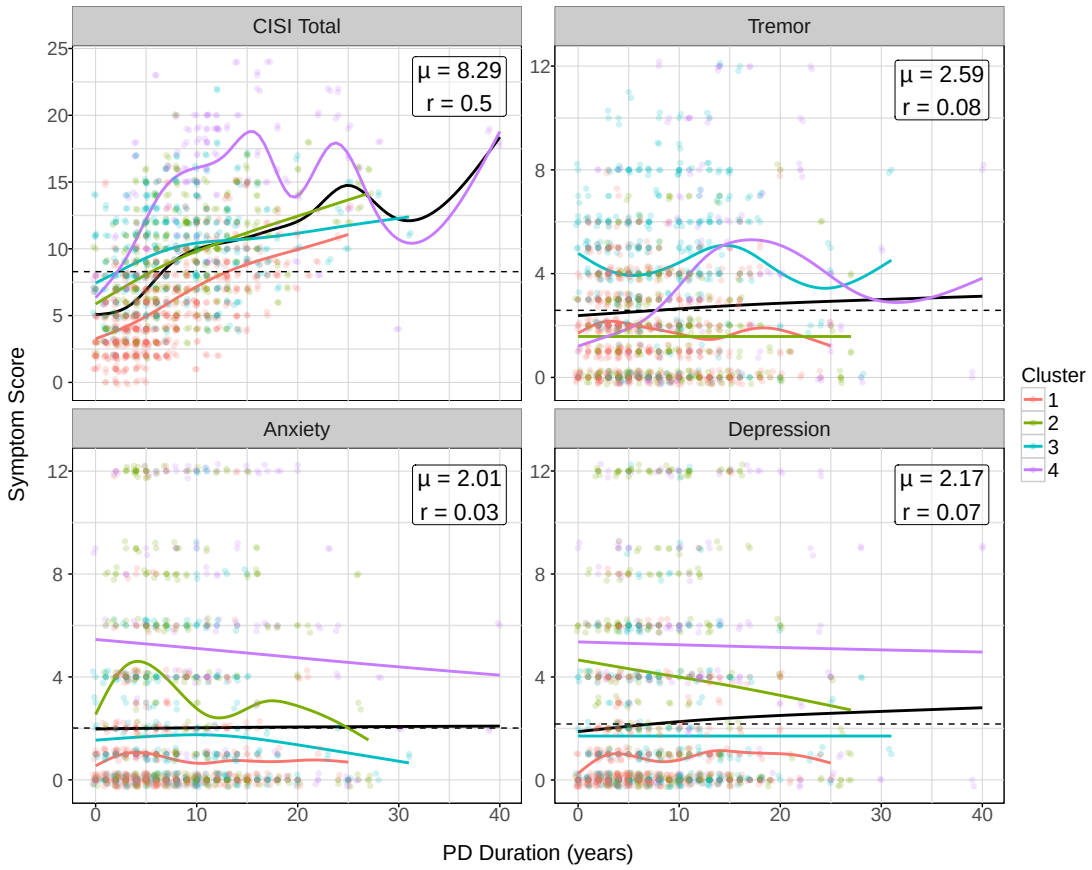

Figure 8: Selected symptoms for 901 patients plotted against PD duration. Smoothed loess curves for each cluster are drawn in their respective colors; the black curve is the curve for the entire population.