Supplementary material 1. Pre-treatment of data prior to clustering.

1.1. Adjustment for differences in field strength and estimated intracranial volume (eTIV):
Thickness and volume measures in the PD and HC groups were adjusted for field strength (1.5T vs 3.0T) using the residual approach (Voevodskaya et al., 2014). Additionally, volume measures were adjusted for eTIV. The HC group (combined across EXPANd, PPMI, Barcelona UB-Clínic) was used as the reference group.

1.2. Adjustment for cohort differences:
Following the adjustments in 1.1, systematic differences persisted when compared across the four cohorts. Thus, when individuals were combined across all cohorts (KIMOVE, EXPANd, PPMI, Barcelona UB-Clínic), thickness and volume measures in both the PD and HC groups were additionally adjusted for cohort using the residual approach (Voevodskaya et al., 2014; Poulakis et al., 2022). The PPMI cohort was used as the reference group due to its largest sample size.

References:
Voevodskaya O, Simmons A, Nordenskjöld R, Kullberg J, Ahlström H, Lind L, Wahlund LO, Larsson EM, Westman E; Alzheimer's Disease Neuroimaging Initiative. (2014). The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. Front Aging Neurosci, 7;6:264.
Poulakis K, Pereira JB, Muehlboeck JS, Wahlund LO, Smedby Ö, Volpe G, Masters CL, Ames D, Niimi Y, Iwatsubo T, Ferreira D, Westman E; Japanese Alzheimer's Disease Neuroimaging Initiative; Australian Imaging, Biomarkers and Lifestyle study. (2022). Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. Nat Commun, 13(1):4566.

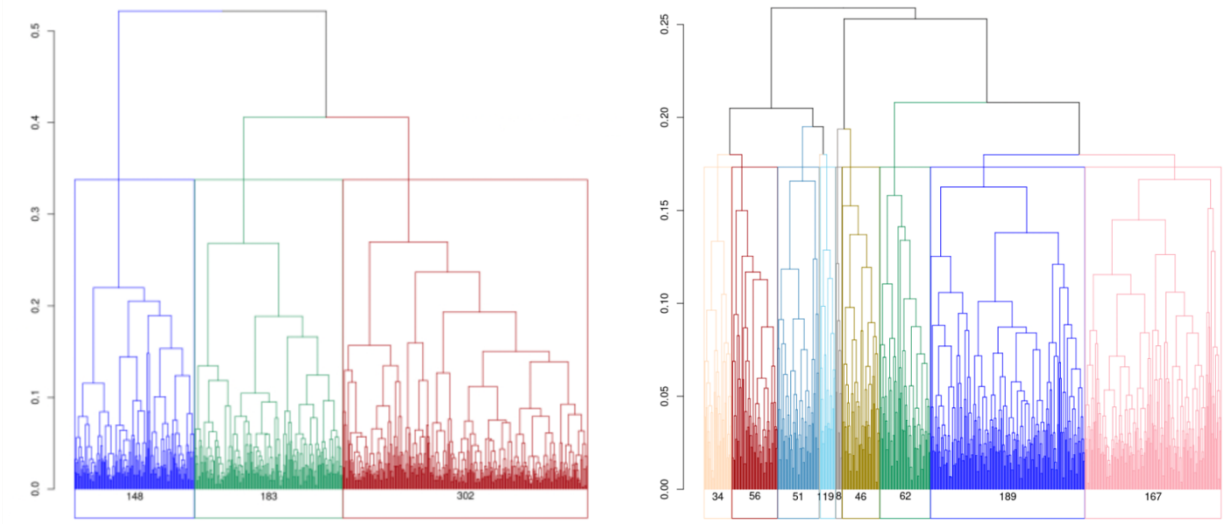
Supplementary material 2. List of indices used to identify the optimal cluster solution.

(**1**) Krzanowski and Lai (KL) index; (**2**) Calinski-Harabasz (CH) index; (**3**) Hartigan index; (**4**) Cubic clustering criterion (CCC); (**5**) Scott index; (**6**) Marriot index; (**7**) Trace within-cluster covariance; (TrCovW) index; (**8**) Trace within-cluster sum of squares (TraceW) index; (**9**) Friedman index; (**10**) Rubin index; (**11**) C-index; (**12**) Davies-Bouldin (DB) index; (**13**) Silhouette index; (**14**) Duda index; (**15**) Pseudot2 index; (**16**) Beale index; (**17**) Ratkowsky index; (**18**) Ball index; (**19**) Point biserial index (Ptbiserial); (**20**) Fraley index; (**21**) McClain index; (**22**) Dunn index; (**23**) Hubert index; (**24**) Standard deviation index (SDindex); (**25**) S-index; (**26**) SFbw index.

Supplementary material 3. Dendrograms from the cluster analysis.

(A) Not adjusted for global atrophy

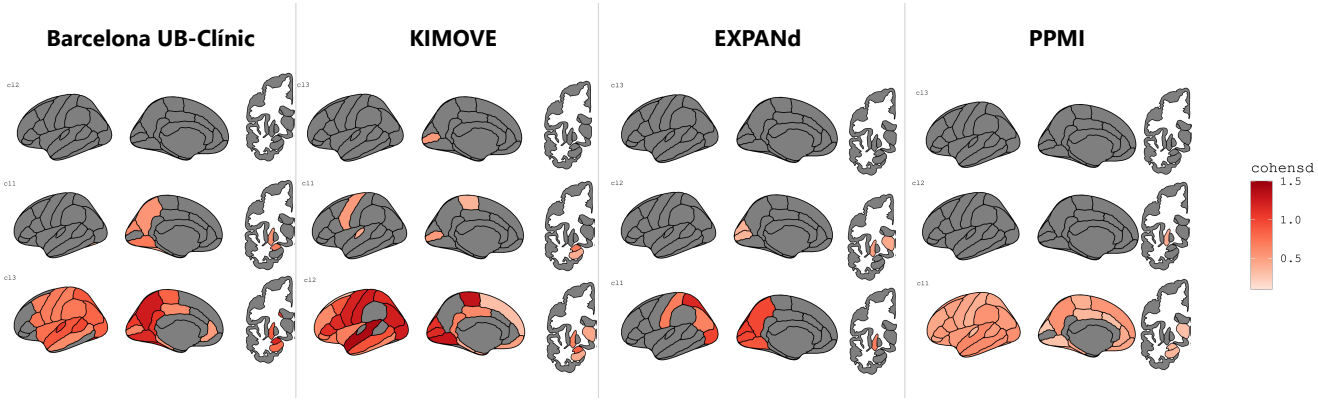
(B) Adjusted for global atrophy



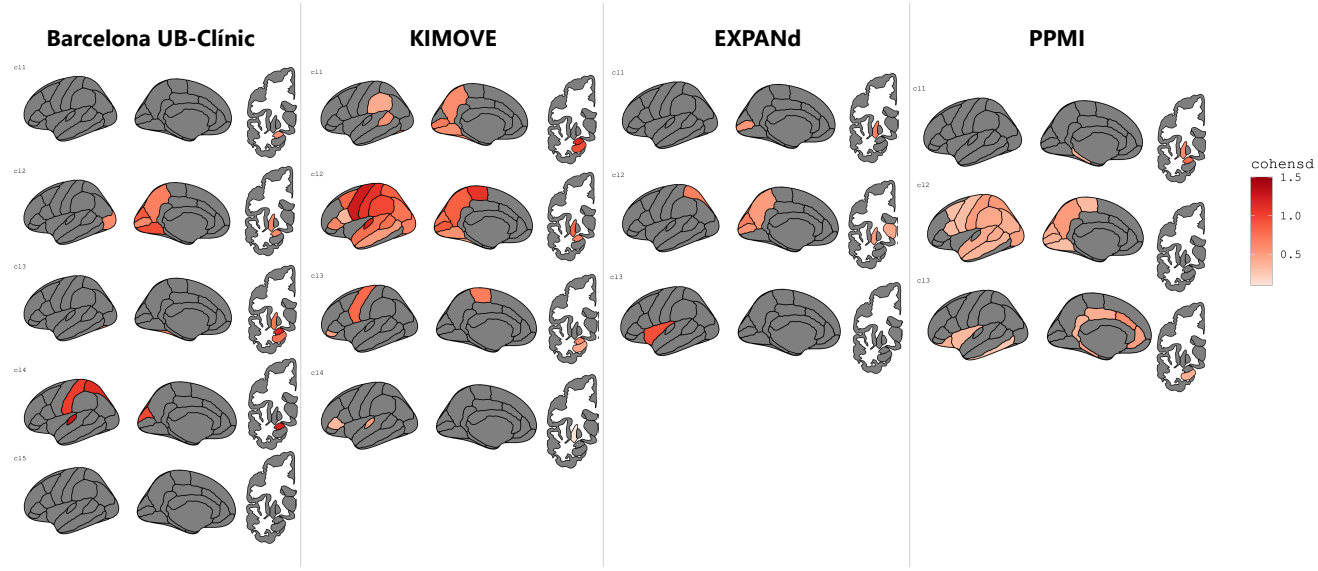
Dendogram or tree structure split and colour-coded based on the optimal cluster solution for each clustering model – (A) 3 cluster solution is shown for clustering without adjustment for global atrophy and (B) 10 cluster solution is shown for clustering with adjustment for global atrophy. The vertical axis shows the inter-cluster distance based on the random forest similarity between individuals.

Supplementary material 4. Brain atrophy patterns of each cohort separately.

(A) Global atrophy-unadjusted

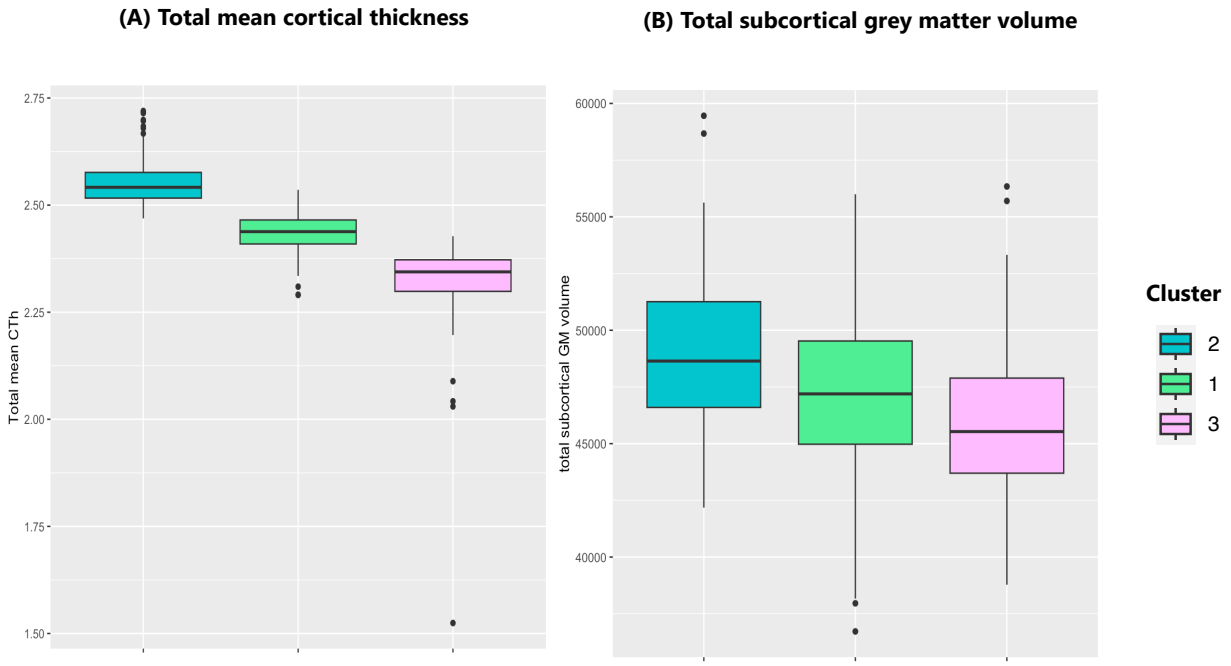


(B) Global atrophy-adjusted



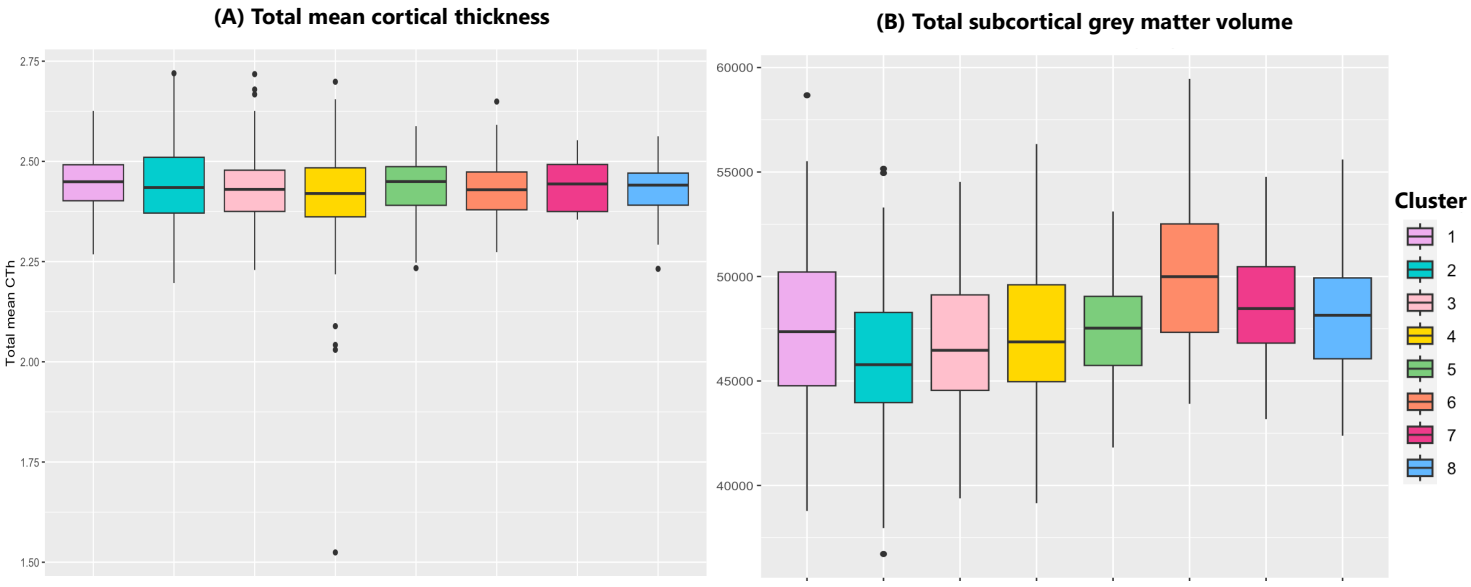
PD brain atrophy patterns compared to the HC group. The maps represent an average across the right and left hemispheres. Cohen's d values were calculated and are presented in the figure when the p-value of the ANCOVA reached <0.05. Darker red indicates more atrophy in the cluster than in HC. Results were adjusted for age.

Supplementary material 5. Overall atrophy measures across the global atrophy-unadjusted clusters.



Boxplots displaying (A) total mean cortical thickness and (B) global mean subcortical grey matter volumes across the 3 PD clusters obtained from global atrophy-unadjusted clustering. All pair-wise comparisons were significant ($p<0.05$).

Supplementary material 6. Overall atrophy measures across the global atrophy-adjusted clusters.



Boxplots displaying (A) total mean cortical thickness and (B) global mean subcortical grey matter volumes across the 8 PD clusters obtained from global atrophy-adjusted clustering. PD subtypes did not show significant differences in total mean cortical thickness (A). However, significant differences ($p<0.05$) in total subcortical grey matter (B) were observed between the following pairwise comparisons: cl2 vs cl4; cl2 vs cl6; cl2 vs cl8; cl6 vs cl1; cl6 vs cl3; cl6 vs cl4; cl6 vs cl5.

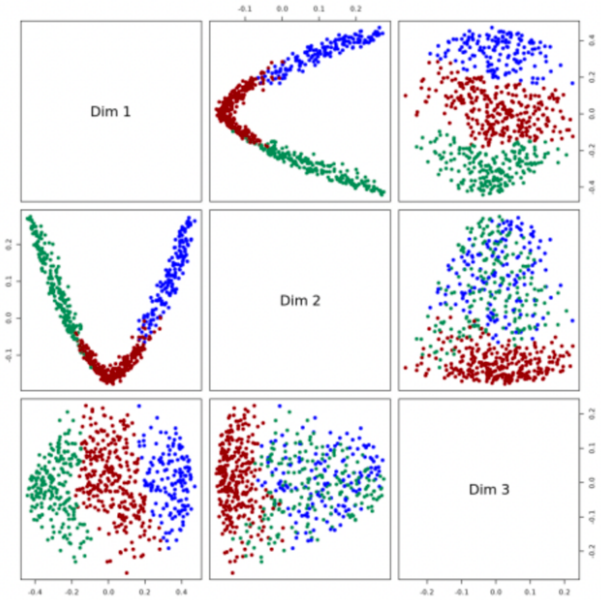
Supplementary material 7. Breakdown of samples contributing to clinical progression models for UPDRS-III and MoCA.

UPDRS-III							
	Baseline visit	1-year follow up	2-year follow up	3-year follow up	4-year follow up	5-year follow up	6-year follow up
PPMI	310	273	280	272	244	217	12
Barcelona UB-Clínic	104	6	3	61	0	40	17
MoCA							
PPMI	310	273	280	272	244	217	12
Barcelona UB-Clínic	104	9	0	0	0	45	22

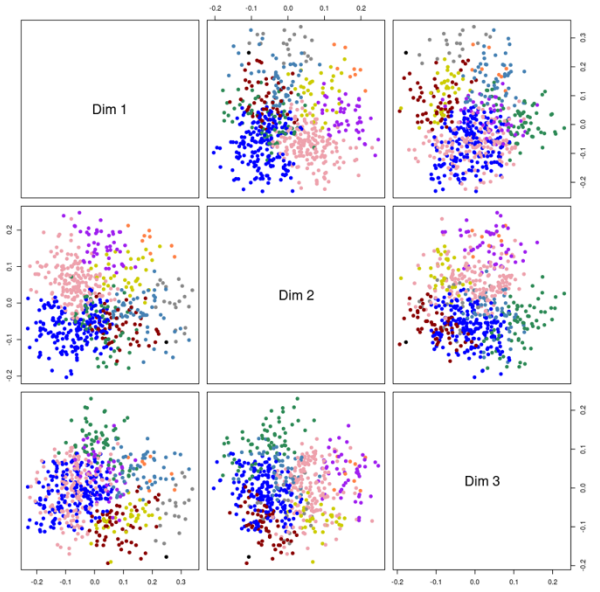
Abbreviations: Barcelona UB-Clínic – Cohort from Hospital Clinic, Barcelona; MoCA – Montreal Cognitive Assessment; PPMI – cohort from the Parkinson's Progression Markers Initiative; UPDRS - Unified Parkinson's Disease Rating Scale.

Supplementary material 8. Multidimensional scaling.

(A) Not adjusted for global atrophy



(B) Adjusted for global atrophy



3D multi-dimensional scaling representation of the similarity matrix. The plots represent the optimal cluster solution for each clustering model – (A) 3 cluster solution is shown for clustering without adjustment for global atrophy and (B) 10 cluster solution is shown for clustering with adjustment for global atrophy. Dots represent PD subjects, the distance between them symbolizes how similar the subjects are with respect to their patterns of brain atrophy, and the colour denotes the clusters obtained by the hierarchical clustering.