Background Conventional “top-down” classification approaches depend on a-priori assumptions about disease phenotype. “Bottom-up” data-driven statistical approaches in well-characterised, incident cohorts are relatively novel and may have specific advantages for disease stratification. The Discovery cohort is a population-based, prospective early PD cohort that aims to better understand the biology and identify prognostic markers through comprehensive motor/non-motor longitudinal assessment.

Methods 454 consecutive patients (UKPDBB probable PD diagnosis ≤3.5 years) were included in factor and k-means cluster analysis which discriminated between patients using motor and non-motor baseline assessments (MDS-UPDRS I-IV, H&Y, Timed walk, Flamingo test, Purdue Pegboard, Sniffin smell, Constipation and RBD Questionnaires, Epworth Sleep Scale, Leeds Anxiety and Depression Scale, MMSE, MoCA, verbal fluencies)

Results 5 patient clusters were identified: Group 1 (28%) were younger and had milder disease, Group 2 (19%) scored poorly on constipation and psychological well-being, Group 3 (18%) were tremor dominant, Group 4 (14%) had markedly poorer psychological well-being, Group 5 (21%) were driven by cognitive/postural problems.

Conclusions This data-driven approach identified novel patient clusters (Groups 2, 4) characterised by constipation and poorer psychological well-being that would have been missed using traditional motor classification. Future work will assess whether group memberships predicts treatment-response and progression, including dementia and motoric decline.